BBC RD 1984/7

RESEARCH DEPARTMENT

REPORT

Video noise reduction

J.O. Drewery, M.A., Ph.D., C.Eng., M.I.E.E.
R. Storey, B.Sc., C.Eng., M.I.E.E.
N.E. Tanton, M.A., C.Eng., M.I.E.E., M.Inst.P

Research Department, Engineering Division
THE BRITISH BROADCASTING CORPORATION

July 1984

VIDEO NOISE REDUCTION

J.O. Drewery, M.A., Ph.D., C.Eng., M.I.E.E.
R. Storey, B.Sc., C.Eng., M.I.E.E.
N.E. Tanton, M.A., C.Eng., M.I.E.E., M.Inst.P

Summary

A video noise and film grain reducer is described which is based on a first-order recursive temporal filter. Filtering of moving detail is avoided by inhibiting recursion in response to the amount of motion in a picture. Motion detection is based on the point-by-point power of the picture difference signal coupled with a knowledge of the noise statistics. A control system measures the noise power and adjusts the working point of the motion detector accordingly. A field trail of a manual version of the equipment at Television Centre indicated that a worthwhile improvement in the quality of noisy or grainy pictures received by the viewer could be obtained. Subsequent trials of the automated version confirmed that the improvement could be maintained. Commercial equipment based on the design is being manufactured and marketed by PYE T.V.T. under licence. It is in regular use on both BBC1 and BBC2 networks.

Issued under the Authority of

M.E.S. Roffel

Head of Research Department

Research Department, Engineering Division
BRITISH BROADCASTING CORPORATION

(PH-256)

July 1984

VIDEO NOISE REDUCTION

J.O. Drewery, M.A., Ph.D., C.Eng., M.I.E.E.
R. Storey, B.Sc., C.Eng., M.I.E.E.
N.E. Tanton, M.A., C.Eng., M.I.E.E., M.Inst.P.

VIDEO NOISE REDUCTION

J.O. Drewery, M.A., Ph.D., C.Eng., M.I.E.E.
R. Storey, B.Sc., C.Eng., M.I.E.E.
N.E. Tanton, M.A., C.Eng., M.I.E.E., M.Inst.P.

1. Introduction

Noise is an inherent feature of any communication system and, in a broadcasting network, arises in many different ways. At the source it may be generated as photon noise or thermal noise in the television camera or as film grain in the cine camera. Circuits used to carry the signal then add further noise and recorders add yet more as the signal effectively recirculates round a noisy loop during post-production. Finally the signal is carried to the transmitter via the distribution network which adds more noise and radiated via a noisy transmission path. The sum total of all this is that the impairment due to noise of programmes currently broadcast varies from imperceptible to marginally acceptable.

At the outset of the research described here, many devices were marketed which claimed to reduce video noise. However, these invariably relied on some assumed spatial property of the picture information, so that noise reduction was accompanied by impairment of spatial resolution. This can be avoided by using picture-to-picture integration but such an operation impairs motion unless elaborate precautions are taken. Image Transform in the USA had, since the late 60's, been successfully reducing noise using picture integration as part of a specialised requirement for transferring NTSC video from tape to film.[1] Their equipment incorporated motion detectors and was based on analogue storage devices, so being somewhat bulky. Nevertheless the spectacular results they achieved with the early NASA space pictures stimulated the work described in this Report in which digital techniques were brought to bear on the problem of developing a similar equipment for broadcast use.

From a system point of view it might be thought that each source of noise should be attacked as it occurs to prevent the noise accumulating to very high values. However, experience with a noise reducer of the type to be described suggests that the small impairments it introduces also accumulate to give an unacceptable result after seven or eight operations. Thus it is, at present, thought that such noise reducers should not be cascaded more than is necessary in the early part of the broadcast chain.

On this basis the logical position of a noise reducer is in the domestic receiver, where it can deal with noise introduced into every part of the signal chain. Although this is by no means impossible, given the current trend in large-scale integrated circuits, current economics dictates that the optimum place for a noise reducer, at least in the BBC's networks, is in the output of the network feeds at Television Centre. In this way only two machines are needed but they must then be flexible enough to deal with the great variety of noise characteristics found in practice.

A further consequence of this approach is that the machine must be capable of dealing with the composite PAL colour signal. In doing this it may be designed either to decode the composite signal into its luminance and colour difference components or to avoid decoding and recoding and make due allowance for the inherent field-to-field difference of the signal caused by the subcarrier offset from picture frequency.

2. The basic concept

Noise can be reduced in video signals without impairing spatial resolution by using the fact that in stationary pictures it causes the only difference between signals obtained in successive scans. Thus, averaging the signals corresponding to successive pictures leads to a noise reduction since the picture-to-picture difference is random with zero mean. Such an averaging amounts to a temporal low-pass filter.

There are two basic ways of implementing such a low-pass filter. Firstly the signal may be passed through a series of delay elements, each of length one picture period, and the signals at each stage added together as shown in Fig. 1. Such a filter is a transversal type and the noise reduction

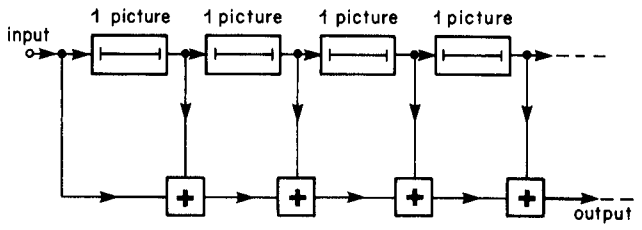


Fig. 1 – Basic transversal filter for noise reduction.

factor, given by the ratio of input to output noise powers, is simply the number of contributions. Thus many picture delay elements would be needed to obtain a large noise reduction. This is the arrangement used by Image Transform, mentioned earlier.

Alternatively, a recursive filter, as shown in Fig. 2, can be used which requires only one picture delay element. The amount of noise reduction is

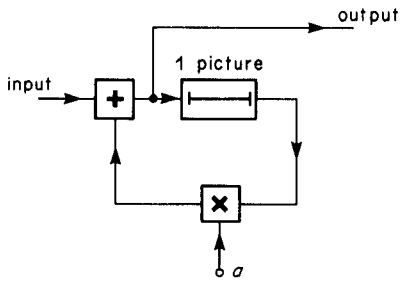


Fig. 2 – First-order recursive filter.

controlled by the feedback factor, a, and, to avoid gain at zero frequency, the filter may be rearranged as in Fig. 3. Here the effect of the subtractor, divider and adder is to form the output as a fraction 1/K of the input picture and a fraction $1 - (1/K)$ of the previous output picture. The impulse response of the filter is thus a decaying exponential sampled at the picture frequency and the effect on moving pictures is very like that of a long persistence display tube with a time constant of K picture periods. With this filter it is possible to

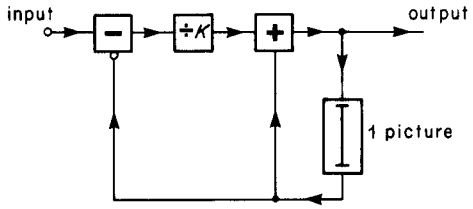


Fig. 3 – Rearrangement of the first-order recursive filter to give unity gain at zero frequency.

obtain very high values of noise reduction factor simply by increasing K but the economy in having only one picture delay is partly offset by the need for an extra bit of accuracy each time K doubles, assuming the filter is realised digitally. The behaviour of this type of filter and its effect on white noise and moving areas is dealt with in a previous Research Department Report.[2]

Fig. 4 shows the frequency characteristic of the filter for a particular value of K. Because the delay element is one picture period the characteristic repeats at the picture frequency T_P^{-1}. The unweighted noise reduction factor is simply the ratio of the integral under the square of the characteristic to that under a flat, unity characteristic, taken over one cycle of the characteristic. This factor can be shown to be given by

$$NF = 2K - 1$$

and is the reduction factor that would be registered by a measuring device with a time constant much longer than the picture period. However, as the reduced noise has the temporal spectral characteristic of the filter it may not appear to the eye to be reduced by the same factor.

To test this hypothesis an experiment was carried out to determine the relationship between the unweighted noise reduction factor given above and the subjective assessment. White noise, added

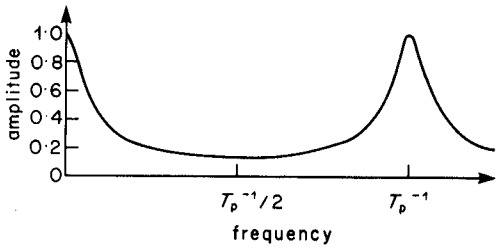


Fig. 4 – The amplitude/frequency characteristic of the recursive filter.

K	Unweighted noise-power reduction $10\log_{10}(2K-1)$ dB*	Subjective noise-power reduction†
2	4.8 dB	4.9 dB
4	8.5 dB	8.1 dB
8	11.8 dB	10.2 dB

* As would be measured by a full-band noise meter.
† Mean values over 6 observers; each observer was asked to match the two halves of a picture; one half had added white noise and was noise-reduced using the linear filter, the other half of the picture had a controllably smaller amount of added white noise and was unprocessed.

Table 1 : Noise-Reduction for Temporal Linear Filtering

to a suitable grey level, which had been passed through a recursive filter of the type shown in Fig. 3 was displayed on one half of a picture monitor. Observers were asked to match this with the unfiltered signal displayed on the other half by varying the attenuation of the added noise. The required attenuation was taken as a measure of the subjective noise reduction and Table 1 shows the results. It is surprising that both the unweighted and the subjective noise reductions differ so little in view of the very disparate nature of the spectra.

In practice it is found that a K value of 4, giving an unweighted noise reduction factor of 8.5 dB, is sufficient for the majority of programme material.

3. Movement protection

3.1. Introduction

As the filter of Fig. 3 behaves rather like a display with a long-persistence phosphor its effect on motion is to cause lag-like smearing of moving objects. This is normally completely unacceptable and so the filtering action must be inhibited wherever motion occurs in the picture. This implies that the noise reappears in moving areas but it is found, in practice, that this is acceptable in the majority of cases because the eye is distracted by the motion; the exception occurs where there is an isolated moving edge. The way in which the filtering action is inhibited is, however, crucial, for if it is done badly the noise-reduced picture may appear worse than the original. The following sections describe the evolution of the movement detection and protection method that is used in the current noise reduction equipment.

3.2. Simple point-by-point methods

A simple method of distinguishing between motion and noise assumes that large interpicture differences are the result of motion and small differences are the result of noise. Consider the filter of Fig. 5 where the simple divider of Fig. 3 has been replaced by a non-linear function F. The

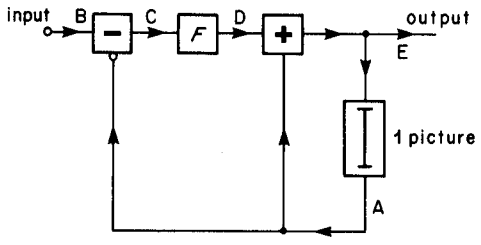


Fig. 5 – The inclusion of a non-linearity in the recursive filter.

transfer characteristic of F is shown in Fig. 6. Inputs to the function F which are smaller in magnitude than a threshold value V_t are divided by K ($K > 1$) whilst inputs larger than V_t are unaltered (corresponding to $K = 1$). Signals which give small interpicture differences (i.e. that are assumed to be noise) are therefore filtered as if by a linear filter (see Ref. 2) whereas those signals giving larger differences (assumed to be moving picture detail) are not filtered at all.

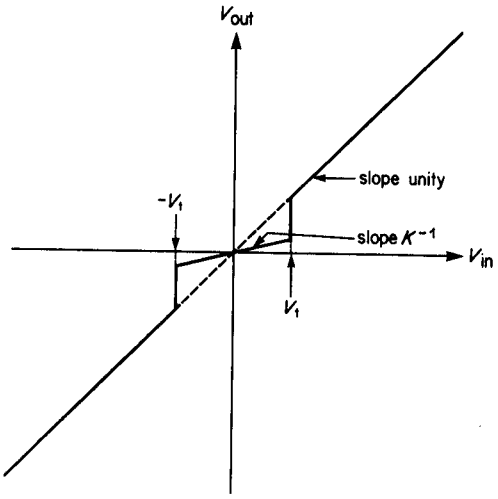


Fig. 6 – Simple discontinuous non-linear function.

This discontinuous non-linear function introduces several picture impairments. Firstly, a threshold suitable for normal noise levels causes low-level moving detail to be smeared because the detail gives rise to differences less than the threshold, and so is filtered. Conversely noise components which give differences greater than the threshold are incorrectly detected as motion. These peaks in noise are not filtered (in contrast to lower amplitude components) and so are rendered more visible as spikey dots amidst the reduced noise. Raising the threshold to eliminate the spikey noise makes the loss of low-level detail even more objectionable.

Secondly, larger moving transitions are followed by residual decaying fringes at spacings governed by the speed of motion. However large the transition, camera integration always ensures that some of the resulting interpicture difference lies beneath the threshold value. The effect is demonstrated in Fig. 7 which shows video waveforms corresponding to a given line in successive pictures when the picture material is an edge moving from left to right. Trace 1 shows the effect of camera integration without processing. Traces 2 to 4 depict the three successive output picture lines when the edge has been processed using the discontinuous function in the filter of Fig. 5. The

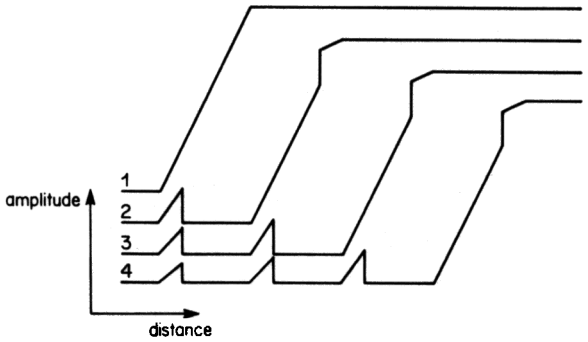


Fig. 7 – The effect on a moving edge of using the discontinuous non-linearity in the recursive filter.

height of each residue decays exponentially with time from an initial value at the beginning of trace 2, dependent on the threshold V_t; the lower the threshold the smaller the residues. Fig. 8 shows this effect on an actual picture. Fringes are clearly visible near the girl's left upper arm.

Thirdly, noise accompanying any moving detail causes vacillations in the threshold decision. Thus an otherwise sharp area of moving detail is broken up into areas which are being filtered and others which are not. The size of these areas is generally of the order of a pixel because it is the high-frequency noise which causes them; the result

Fig. 8. – The fringing effect on a moving picture caused by the mechanism of Fig. 7. Only the upper half is processed.

is an objectionable "grittiness" when movement occurs. Any edge, however large, will produce this effect, again because camera integration ensures that part of the interpicture difference signal is less than the threshold and noise then "dithers" the decision. The inset in Fig. 9 shows a scene impaired by this grittiness.

Fig. 9 – The gritty effect on a moving picture of using the discontinuous non-linearity of Fig. 6. Only the central part inside the white border is processed. The threshold is 9 dB above the r.m.s. value of the noise.

To avoid this problem a function with a continuous characteristic was also studied, see Fig. 10. Interpicture differences less than the threshold V_t are divided by a fixed number K ($K > 1$) as before. However, differences greater than the threshold are just reduced in magnitude by a constant amount in order to keep the transfer characteristic continuous at V_t and to retain unity slope for differences greater than V_t. Grittiness no longer occurs. However, because all interpicture differences are reduced in magnitude (either by division or by subtraction) no moving detail is free from filtering. The effective value of the kernel K never reaches unity however large the difference (although for differences greater than V_t the effective value of K decreases asymptotically towards unity). The result is a loss of resolution on all moving detail and exponential trails following

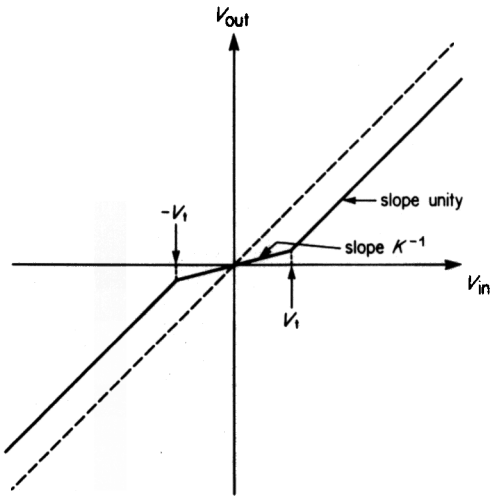


Fig. 10 – Simple continuous non-linear function.

all movement. Fig. 11 shows a scene, the inset part of which has been processed in this way.

3.3. Spatially correlated methods

The obvious criticism of the point-by-point method is that decisions about movement are totally uncorrelated from pixel to pixel whereas, in reality, movement decisions would be highly spatially correlated because objects do not break up into fragments. Thus there is a need to smooth the movement decision. Three ways of doing this were tried and will be described here.

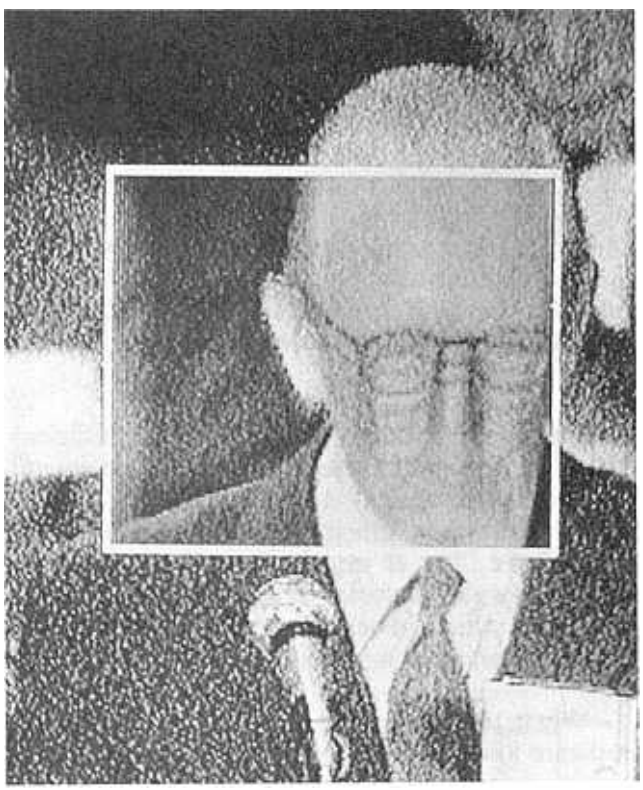


Fig. 11 – The smearing effect on a moving picture of using the continuous non-linearity of Fig. 10. Conditions are as in Fig. 9.

3.3.1. Low-pass filtering

Movement of edges in the picture generates inter-picture differences which are pulses of width equal to the distance travelled in one picture. These pulses are superimposed on the background of unwanted noise and so the detection of moving edges reduces to a matched filter problem where the pulse response of the matched filter is the same as the pulses to be detected, assuming the noise is white. Clearly a compromise must be struck on the assumed pulse width as the speed of motion varies widely from zero to upwards of 30 pixels per picture.

This idea of matched filtering can be extended to two dimensions with profit for the picture difference created by a vertical edge moving horizontally has no vertical variation as shown in Fig. 12. Thus, averaging the picture difference of many adjacent scanning lines provides a further detection improvement. Similarly the horizontal averaging of differences created by a horizontal edge moving vertically gives a "free" detection improvement. So, in general, a two dimensional filter with the same pulse response in the horizontal and vertical directions is required.

To test the hypothesis that the introduction of such a filter would yield a worthwhile improvement the behaviour of the entire digital recursive filter shown in Fig. 13 was simulated using a computer program. The overall circuit is shown at (a) from which it can be seen that the non-linearity, F, of Fig. 5 has been replaced by a side-chain non-linearity, G, in conjunction with a multiplier. A filter, H, precedes the non-linearity and details of the filter, H, and non-linearity, G, are shown at (b) and (c). The number of bits at each point in the overall filter and the parameters of H and G could be independently controlled.

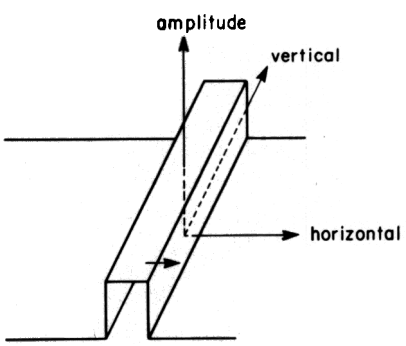


Fig 12 – Two-dimensional representation of the picture difference caused by a moving edge.

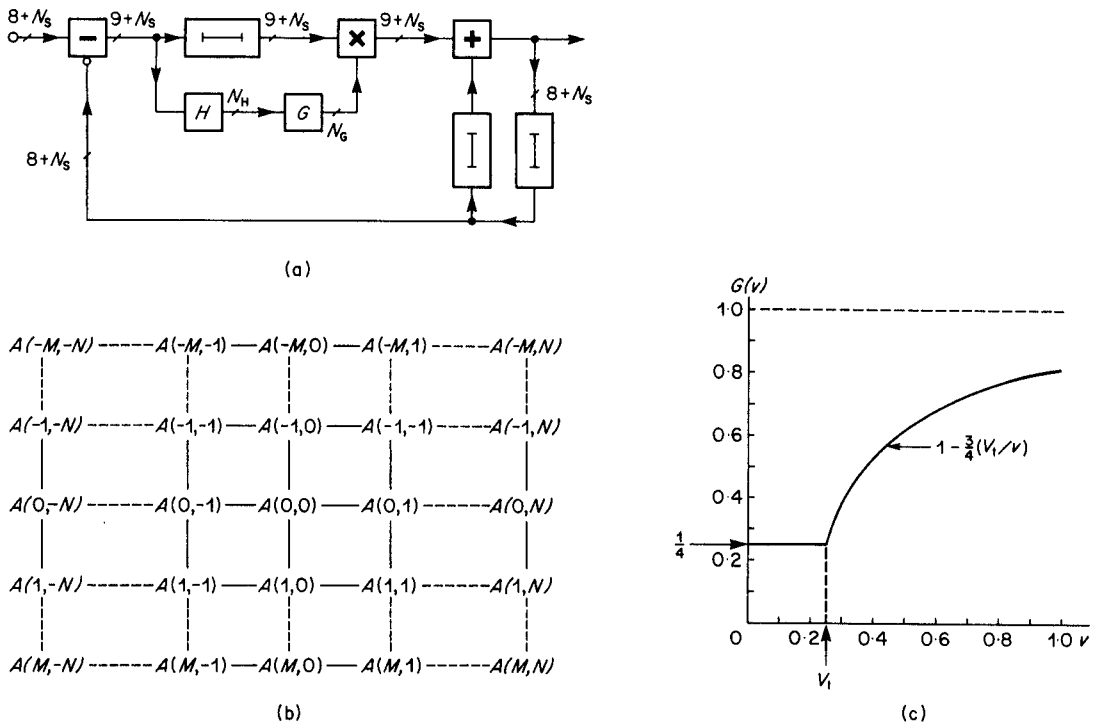


Fig. 13 – The filter simulated by computer, (a) overall circuit (b) detail of filter, H (c) detail of non-linearity, G.

The behaviour was simulated using an input signal which consisted of a $2T$ vertical edge (i.e. a sharp edge passed through a $2T$, i.e. a $\sin^2$, filter) moving from left to right at a given speed, with added white noise derived from a random number generator. The results were plotted in the form of the video signal corresponding to a particular line in successive pictures; by selecting a suitable line care was taken to avoid edge effects when the filter H was two-dimensional.

Some examples of results are shown in Fig. 14. In all cases except (g) and (h) the edge height is 0.7 V and the noise level is 30 dB below the pk–pk signal level of 1.234 V, i.e. about −25 dB relative to 0.7 V. The traces of successive scans have been displaced vertically by 0.1 V for clarity with the last trace at the bottom. Note how the noise level settles to an equilibrium value after two or three scans, caused by starting up the filter with a cleared picture store.

Figs. 14(a), (b) and (c) show the effect of varying the threshold V_t with no filter H. The threshold values are 9, 5 and 3 dB respectively above the r.m.s. noise level. The exponential trails are clearly visible, progressively reducing, coupled with slightly increasing residual noise. Choosing the 5 dB value of threshold Fig. 14(d) shows the effect of introducing a 25 equi-value term filter at H and reducing the threshold by a factor of 25 i.e. 14 dB; the residual noise is almost as small as (b) but the exponential trail has been markedly curtailed. Fig. 14(e) shows the effect of replacing the 25 term one-dimensional filter by a 5 × 5 term two-dimensional filter. The residual noise level is even lower with a further improvement in the exponential trail. Fig. 14(f) shows the same conditions for a faster moving edge. Again no significant trails are evident. To test the effect of non-linearity in the movement detection process, Fig. 14(g) shows the same conditions as in Fig. 14(d) for an edge of only 0.1 V amplitude. As it is difficult to discern the edge with such a low signal-to-noise ratio (8 dB) Fig. 14(h) shows the same conditions but with the input noise removed; as can be seen there is little in the way of an exponential trail.

These simulations were taken as sufficient indication that the inclusion of a two-dimensional filter in the movement detector was very worthwhile. This was a crucial decision as the group delay of the filter H largely determines the delay of the movement detector which must be compensated and which affects the group delay of the whole recursive arrangement.

When such an arrangement was realised in hardware and used to process video signals in real time its defects became apparent. Firstly moving edges were smeared more than predicted and low-level texture was smeared. Secondly, certain pure spatial frequencies, such as occur on test patterns,

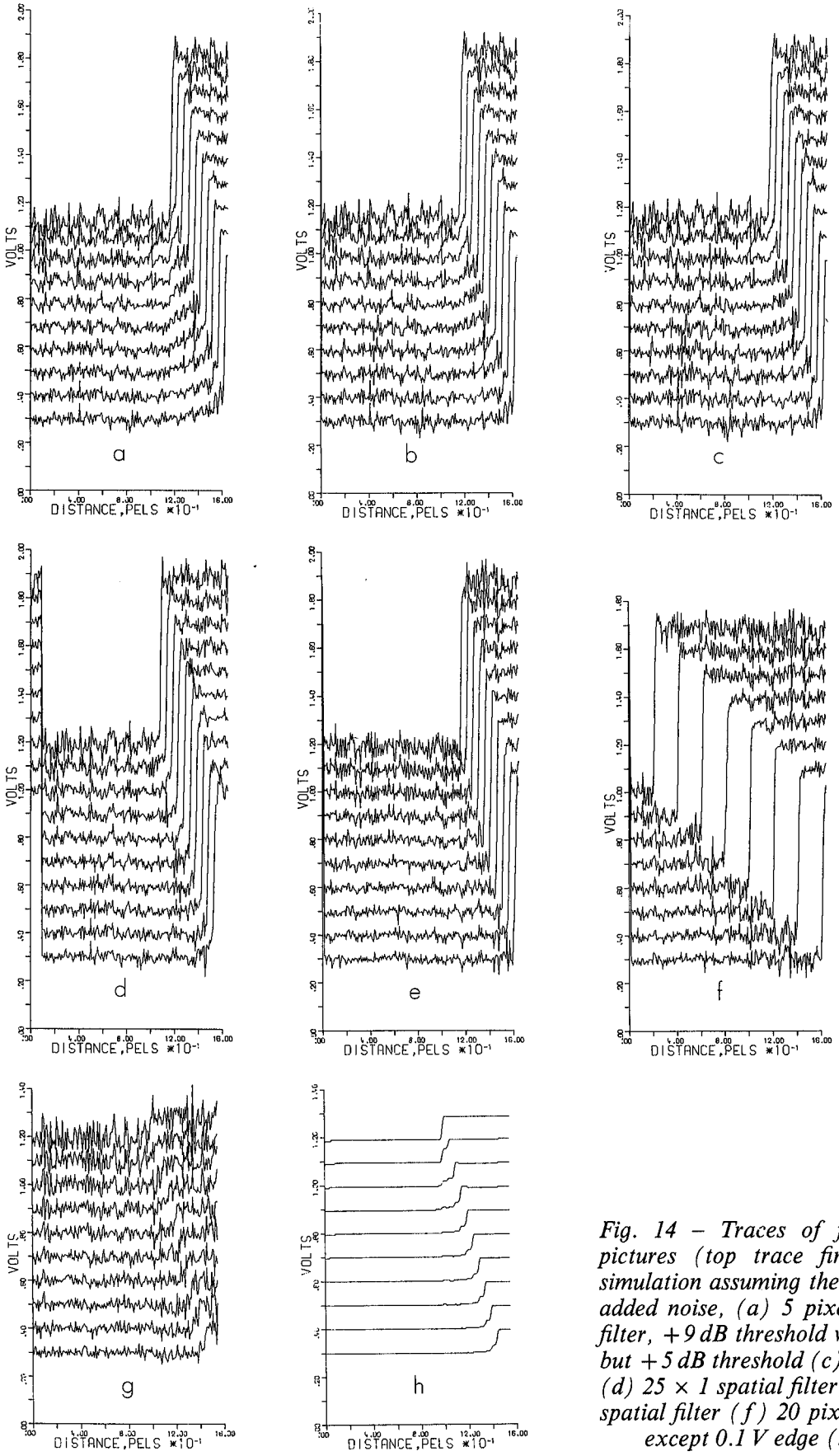


Fig. 14 – Traces of filter output for successive pictures (top trace first) obtained by computer simulation assuming the input is a moving edge with added noise, (a) 5 pixels/picture speed, no spatial filter, +9 dB threshold w.r.t. r.m.s. noise (b) as (a) but +5 dB threshold (c) as (a) but +3 dB threshold (d) 25 × 1 spatial filter, −9 dB threshold (e) 5 × 5 spatial filter (f) 20 pixels/picture speed (g) as (d) except 0.1 V edge (h) as (g) without noise.

developed undesirable effects when moving. This last effect was thought to be caused by the frequencies falling at one of the two-dimensional nulls of H making them completely invisible to the non-linearity and thus invoking a large degree of filtering, fluctuating as the frequency varied slightly.

The first effect was first thought to be caused by camera integration which was not taken into account in the computer simulation. Integration causes a moving edge to be turned into a ramp as in Fig. 7 and causes the picture difference signal to become trapezoidal. However, such an effect is small after the signal has passed through the filter H unless the picture-to-picture displacement is of the same order as the filter width, as shown in Fig. 15. More importantly, the filter H obliterates differences caused by moving texture so making its detection impossible.

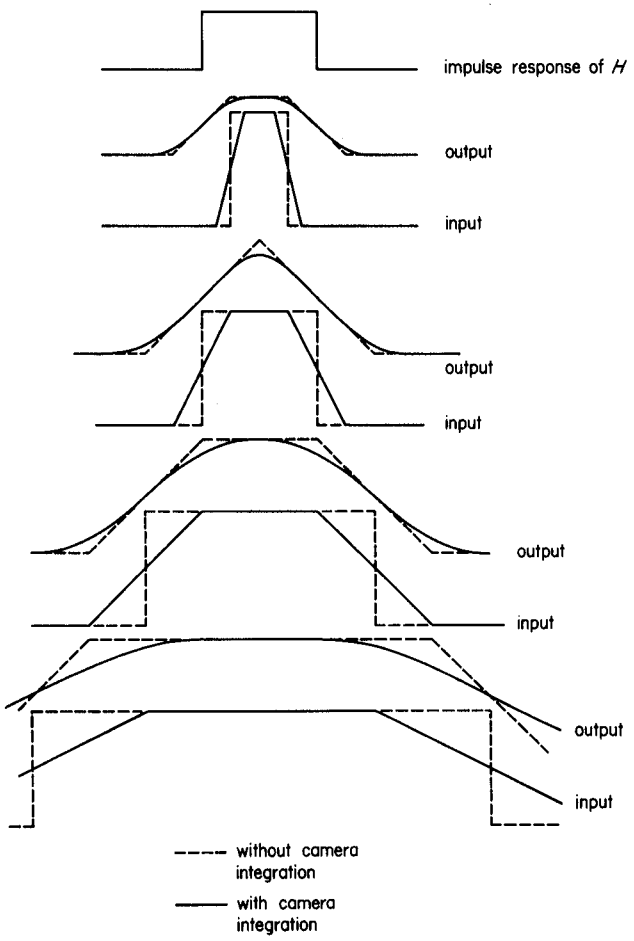


Fig. 15 – The effect of camera integration on the input and output of the spatial filter for increasing widths of input pulse.

3.3.2. Decision spreading and clustering

For the original point-by-point methods of section 3.2 it was pointed out that camera integration was responsible for generating the fringes, shown in Fig. 7, when the discontinuous non-linearity was used. The mechanism of generation is shown in Fig. 16 which shows waveforms at different points in the filter of Fig. 5 for three successive pictures when the input signal is a moving edge, allowing for camera integration. It can be seen that the fringes result from the thresholded signal's being unable to cover the entire region occupied by the picture-difference signal. So the effect can be avoided by spreading the thresholded signal sufficiently depending on the speed of motion. Without knowledge of the motion a compromise amount of spread must be used and the spreading must be two-dimensional to encompass all directions of motion. An advantage is that the dithering decision behaviour mentioned in Section 3.2 is eliminated. The drawback is that a moving edge will be surrounded by an outline of noise because noise reduction will be inhibited in the area of the spread decision.

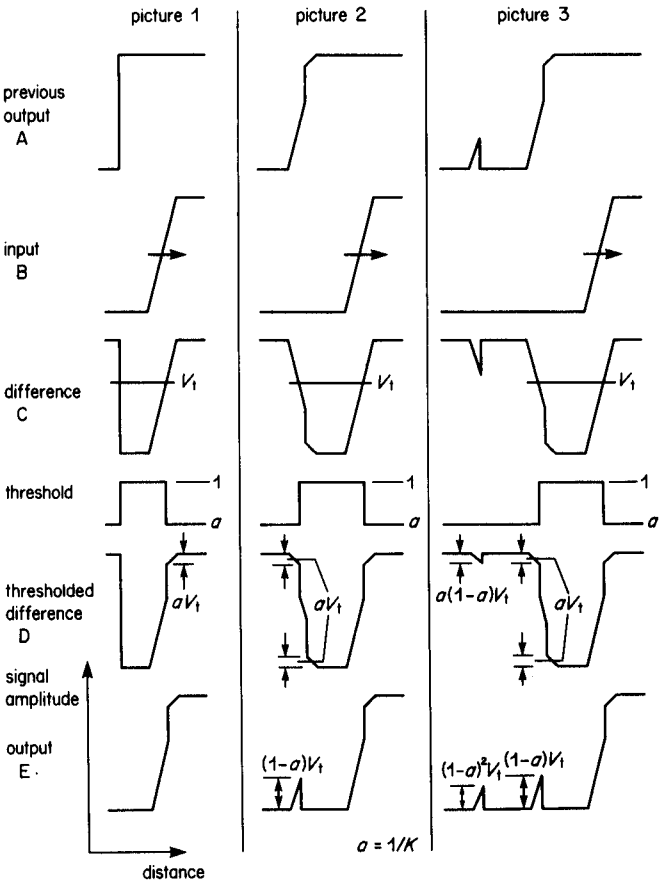


Fig. 16 – The mechanism of generation of fringes when using the discontinuous non-linearity.

This technique was studied using a variables-separable approach (independent horizontal and vertical spread) which resulted in a rectangular area of spread with the movement decision point at its centre. Having chosen an appropriate degree of spreading the fringes were eliminated and the "grittiness" associated with the discontinuous non-linearity disappeared. However, in stationary areas noise peaks which were incorrectly detected as motion were replaced by areas of unprocessed (i.e. noisy) picture, having the size of the spreading aperture. This appeared somewhat like falling raindrops and the upper half of the off-screen photo, in Fig. 17 shows the effect on a plain grey area. The decision process also vacillated because noise on moving edges dithered the movement detection. This resulted in rectangular areas of picture near moving edges being filtered and then not filtered, a process which broke up the texture of the edge. The upper half of Fig. 18 attempts to show the effect.

To reduce the visibility of these effects a graded spreading was studied. The sharp filtering decision was replaced by a gradual stepped transition from no filtering at the point of decision to full filtering at several pixels away from the point. Fig. 19 shows diagrammatically the profile of the graded spreading studied.

Fig. 17 – The "falling raindrops" effect of spreading movement decisions. Only the upper half is processed.

Though the visibility of raindrops (the areas of unprocessed picture in plain areas) was greatly reduced, because the outlines of spread decisions were no longer sharp, fringes were not completely suppressed and the juddery movement caused by the breaking up of edge texture was still very evident. Graded spreading was not therefore very successful as an improvement over normal spreading.

Although spreading successfully removed trails and most of the fringes it unfortunately increased the area of picture over which noise was not reduced when it was falsely detected as motion. The probability of false detection could be lessened by increasing the threshold but at the expense of filtering out moving low level detail. Fig. 20 shows the probability, P, that a picture difference of magnitude V exceeds the bipolar threshold $\pm V_T$ in the presence of Gaussian noise of r.m.s. value σ. For thresholds greater than $\sqrt{2}$ times the r.m.s. noise level, i.e. $y_t > 1$, the signal level giving 50% probability of detection is equal to the threshold, i.e. $y_t \approx x_V$, whereas the probability of false detection, i.e. for $x_V = 0$, varies very rapidly with threshold. For example, a threshold of $\sqrt{2}$ times the r.m.s. noise level ($y_t = 1$) gives a false detection probability of 15% whereas merely doubling it reduces this to 0.5% whilst only doubling the amplitude of the "50% detectable" detail.

These statistics apply to a uniform picture difference such as might be produced by a sudden change of illumination on a featureless background. On the other hand a moving edge generates a picture difference pulse as previously described in which case the statistics apply to different parts of the pulse. But moving texture generates a picture difference which is more random in character and so adds to the noise on a power basis. Thus the statistics for this kind of signal are somewhat different as can be seen from Fig. 21 where the r.m.s. texture amplitude is σ_t. Note that the rate of increase of detection probability with signal amplitude falls off as the amplitude increases.

When spreading was invoked a way had to be found of lowering the probability of false detection at a given point without sacrificing the movement performance, i.e. of increasing the slope of the curves of Figs. 20 and 21. One way of doing this was to use the observation that, with a threshold giving an acceptable proportion of picture which was not noise-reduced (without spreading), false detections nearly always occurred in isolated pixels whereas true motion resulted in contiguous detec-

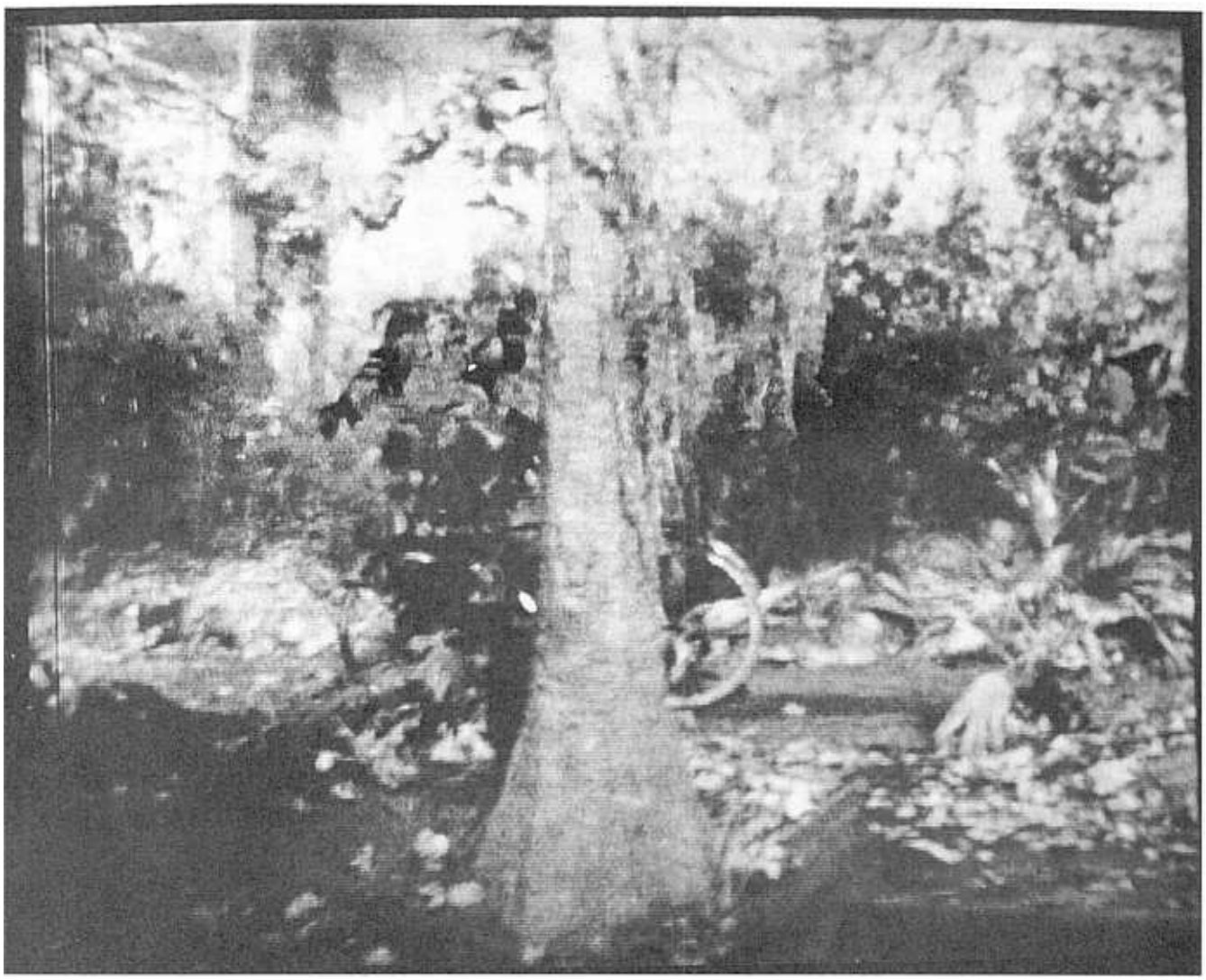


Fig. 18 – The vacillation effect of spreading movement decisions. Note the coarse block pixellation in the upper half of the picture.

tions. Thus motion detection could be based on the idea of simultaneous indications of movement in adjacent pixels, an idea which came to be known as "clustering".

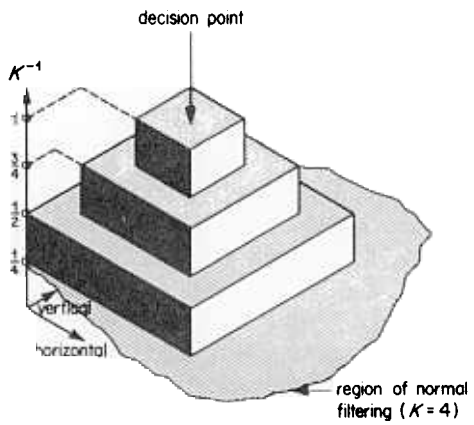
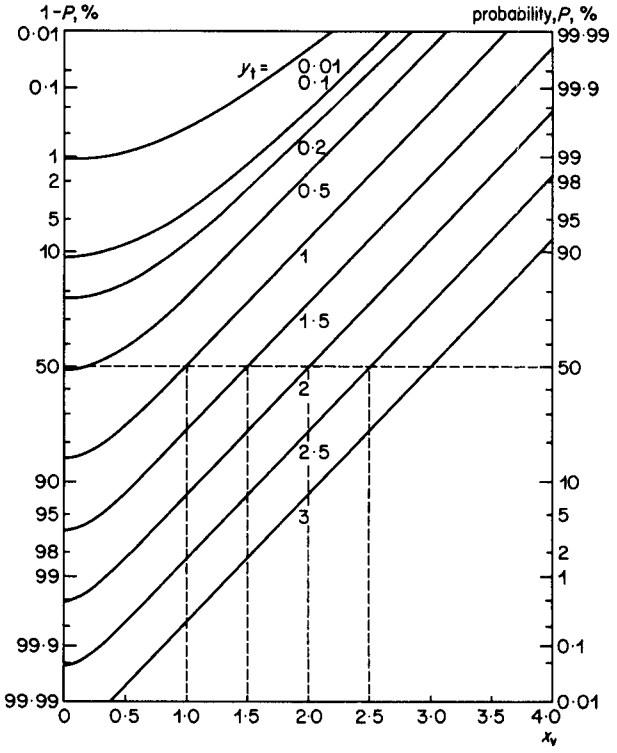
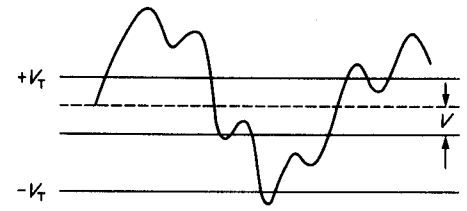


Fig. 19 – The profile of K value with graded spreading.

The size of the cluster could be expected to have an important bearing on the efficacy of the method for if the probability of isolated detections is P then the probability of a "clustered" detection is P^n where the cluster contains n pixels. Fig. 22 shows the theoretical probability, P_n, of such a clustered detection of a uniform picture difference, V, in the presence of Gaussian noise of r.m.s. value σ, for various cluster sizes. The threshold has been set to give a false detection probability (i.e. for $x_V = 0$, in the absence of signal) of 0.1%. To keep the false detection probability constant as n increases P must also increase (since $P < 1$) and therefore the threshold must decrease. It will be noted that although "clustering" does steepen the curve as desired, most benefit is gained from quite modest cluster sizes and there is a limiting case. Even for



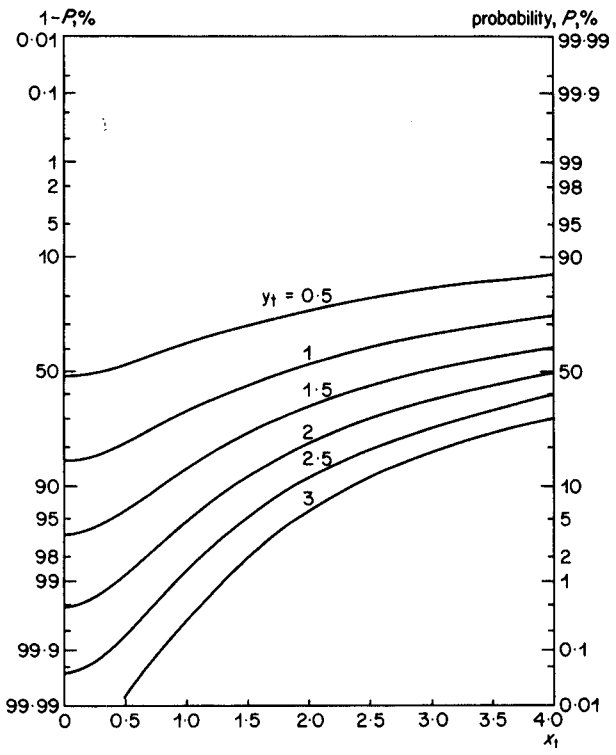


$$P = 1 - \tfrac{1}{2}[\mathrm{erf}(y_t - x_V) + \mathrm{erf}(y_t + x_V)]$$

where

$$y_t = V_T/\sigma\sqrt{2} \quad \text{and} \quad x_V = V/\sigma\sqrt{2}.$$

Fig. 20 – The probability of a uniform picture difference, V, exceeding a threshold, V_T, in the presence of Gaussian noise of r.m.s. value σ.



$$P = 1 - \mathrm{erf}[y_t/(1 + x_t^2)^{1/2}]$$

where

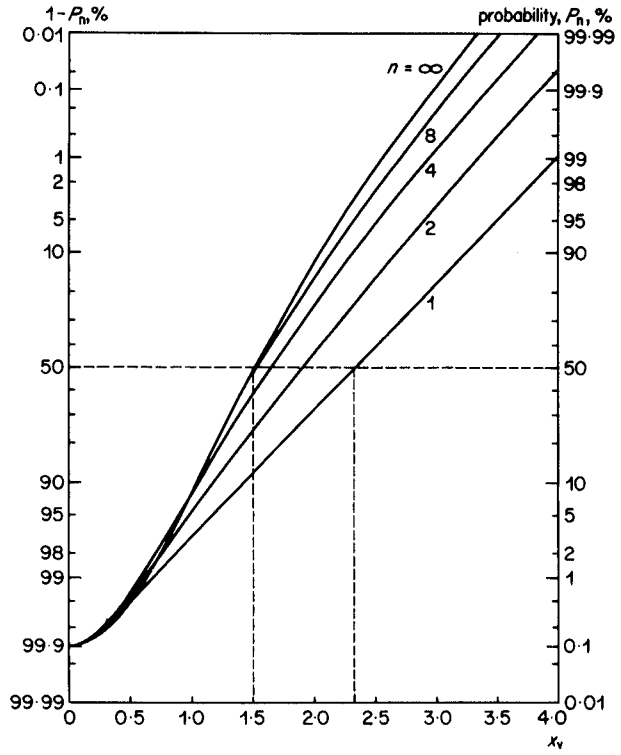
$$y_t = V_T/\sigma\sqrt{2} \quad \text{and} \quad x_t = \sigma_t/\sigma.$$

Fig. 21 – The probability of a random picture difference of r.m.s. value σ_t, such as caused by moving texture, exceeding a threshold in the presence of Gaussian noise.

this limiting case it can be seen, from Fig. 22, that, for a given movement detection probability, the amplitude of the difference signal ($\propto x_V$) is reduced by only about 38% compared with no clustering for the same probability of detection.

Such a scheme, based on clusters of 9, was instrumented and used in conjunction with the spreading technique. When processing real pictures no detection improvement could be discerned although the assessment was probably dominated by the effects of the spreading, described earlier. Moreover the effect of the clustering technique on texture is somewhat different as shown in Fig. 23. As can be seen, clustering actually *reduces* the detectable level of signal, if random. Therefore, the clustering technique was abandoned.

A technique which was not tried experimentally but which received considerable attention in the literature[3] was that of "majority decision" detection. As its name implies, a positive indication at a pixel is registered if motion is indicated at a majority of the pixels in the neighbourhood. Fig. 24 shows the theoretical probability, P_d, of detecting a uniform picture difference with such a detector based on clusters of 8. Curves are shown for majorities of not less than 4, 6, 7 and 8. The last is, of course, the same as the appropriate curve in Fig. 22. It can be seen that a majority of 4 or more gives the most sensitive results and, compared with Fig. 22, gives more than a 50% reduction in detectable signal compared with no clustering. However, the effect of such a detector on texture is shown in Fig. 25 where it will be seen that the most sensitive result is given by choosing a "majority" of 1 or more out of 8. Clearly, then, a true majority decision detector would not be optimum for texture and so this approach to movement protection was also dropped.



$$P_n = P^n$$

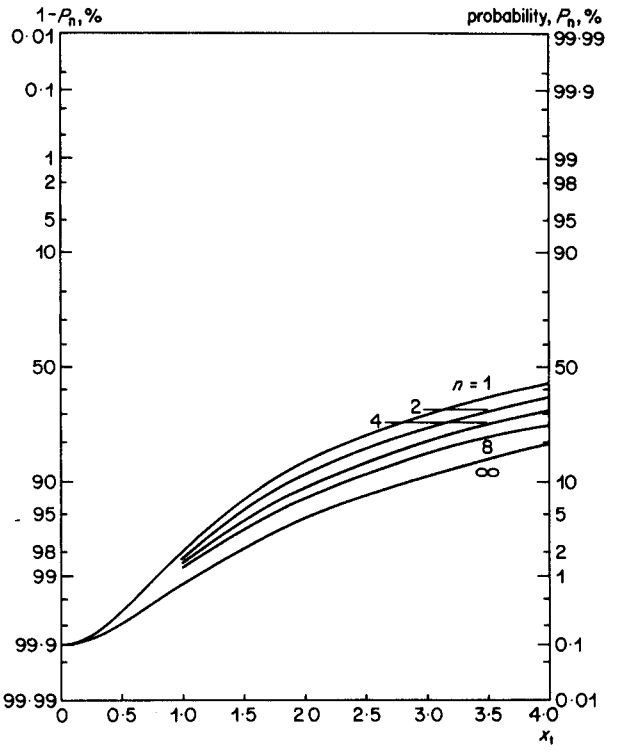
where

$$P = 1 - \tfrac{1}{2}[\mathrm{erf}(y_t - x_v) + \mathrm{erf}(y_t + x_v)]$$

where

$$(1 - \mathrm{erf}\, y_t)^n = 10^{-3}.$$

Fig. 22 – The probability of a uniform picture difference exceeding a threshold simultaneously at a cluster of n points in the presence of Gaussian noise. The threshold is adjusted with cluster size to give a constant probability for zero signal.



$$P_n = P^n$$

where

$$P = 1 - \mathrm{erf}[y_t/(1 + x_t^2)^{1/2}]$$

where

$$(1 - \mathrm{erf}\, y_t)^n = 10^{-3}.$$

Fig. 23 – The probability of a random picture difference exceeding a threshold for the same conditions as in Fig. 22.

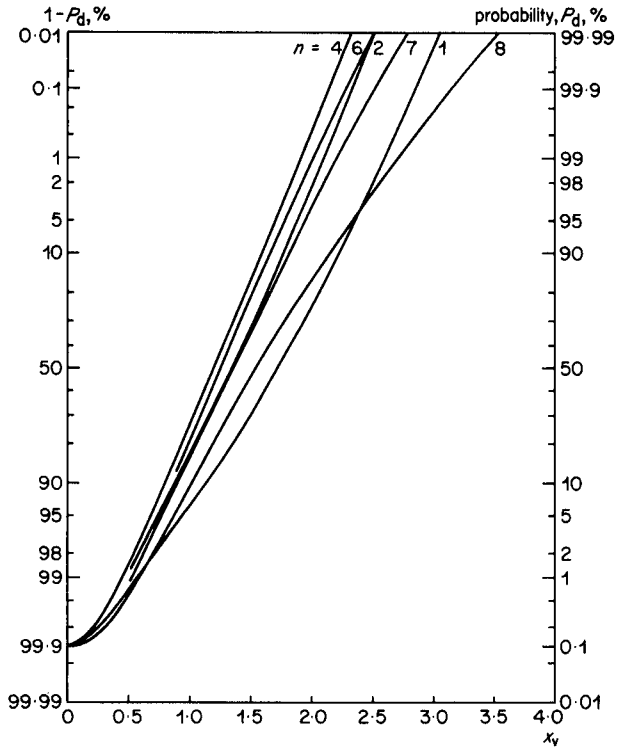
3.3.3. The actual method used

3.3.3.1. Introduction

The criticism of the foregoing cluster and spread method is that its operation is binary in the sense that movement is either detected or not with appropriate filtering. Graded spreading did not help a great deal either as it was still based on a binary decision. What was needed, therefore, was some kind of smooth control wherein the degree of filtering was a function of the smoothed difference signal. This was provided by the low-pass filter method of Section 3.3.1 but that suffered from the disadvantage of missing certain spatial frequencies falling at low-gain points of the filter characteristic. A way of avoiding this is to precede the filter with a rectifier so that the filter input is always positive, and cannot integrate to zero. This is part of the basis of the method used.

A further important consideration is how the smoothed signal emerging from the filter controls the degree of filtering, i.e. the nature of the non-linear relationship, G, in Fig. 13. Clearly, if a smooth control of filtering is required then the function, G, cannot be discontinuous, but must rise from some basal value to unity in a continuous way. As will be seen, there is a limit to how quickly this may occur.

To ensure correct operation of the non-linearity, G, the input must lie at the point where the value of G just begins to rise, in areas where there is no motion. This input value will be that due solely to noise and so will depend on the level of noise. For consistent operation, therefore, the signal emerging from the filter H must be subjected to a variable degree of attenuation, dependent on the noise amplitude.



$$P_d = \sum_{i=n}^{8} {}^8C_i P_n^i (1 - P_n)^{8-i}$$

where

$$P_n = 1 - \tfrac{1}{2}[\mathrm{erf}(y_n - x_v) + \mathrm{erf}(y_n + x_v)]$$

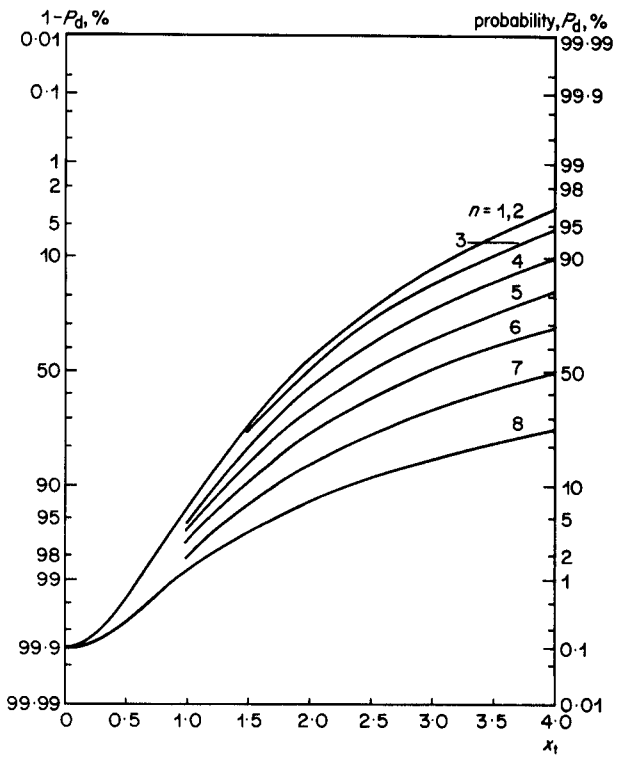
where

$$P_{0n} = 1 - \mathrm{erf}(y_n)$$

where

$$\sum_{i=n}^{8} {}^8C_i P_{0n}^i (1 - P_{0n})^{8-i} = 10^{-3}.$$

Fig. 24 – The probability of a uniform picture difference exceeding a threshold at n or more points in a cluster of 8 in the presence of Gaussian noise. The threshold is adjusted with n to give a constant probability for zero signal.



$$P_d = \sum_{i=n}^{8} {}^8C_i P_n^i (1 - P_n)^{8-i}$$

where

$$P_n = 1 - \mathrm{erf}[y_n/(1 + x_t^2)^{1/2}]$$

where

$$P_{0n} = 1 - \mathrm{erf}(y_n)$$

where

$$\sum_{i=n}^{8} {}^8C_i P_{0n}^i (1 - P_{0n})^{8-i} = 10^{-3}.$$

Fig. 25 – The probability of a random picture difference exceeding a threshold for the same conditions as in Fig. 24.

The four basic elements of the motion detector are thus a rectifier, spatial filter, variable attenuator and a non-linearity and are shown in Fig. 26. These will now be discussed in more detail.

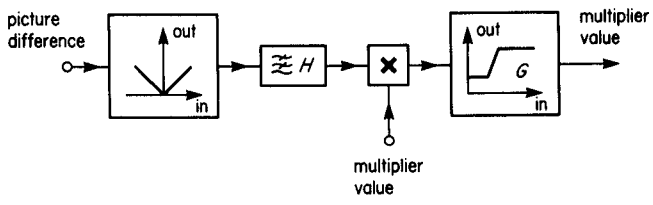


Fig. 26 – The four basic elements of the motion detector.

3.3.3.2. Action of rectifier

The action of the rectifier in conjunction with the spatial filter H is to form the mean modulus of the picture difference signal. This is simpler than but similar to measuring the r.m.s. of the difference signal which is the quantity more usually found in classical solutions to signal processing problems. In the absence of motion this mean modulus is a good representation of the r.m.s. value of the noise which forms the only contribution to the picture difference. Because the filter, H, is finite the mean has a variation and its probability distribution function can be derived as follows.

Assuming the noise has a Gaussian probability distribution, the distribution of the picture difference signal, after rectification, is "half-Gaussian" as shown at Fig. 27(a). Appendix 1 shows that the mean, μ, and variance, v, of this distribution are given by

$$\mu = \sqrt{2/\pi}\,\sigma \qquad v = \sigma'^2 = (1 - 2/\pi)\sigma^2$$
$$\approx 0.8\sigma \qquad\qquad \approx 0.36\sigma^2 \qquad \ldots(1)$$

where σ is the standard deviation of the original Gaussian distribution, i.e. the r.m.s. noise value. Thus the standard deviation of the half-Gaussian distribution, σ', is about $\frac{3}{4}$ of the mean.

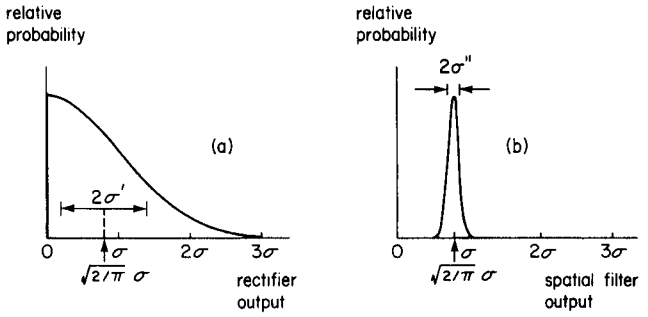


Fig. 27 – The probability distribution of the picture difference signal, assuming Gaussian statistics. (a) after rectification (b) after rectification and spatial filtering.

The spatial filter forms an equal-weight sum of many adjacent pixel differences and so, assuming that the differences are uncorrelated, i.e. the noise is spatially white, the probability distribution of the filter output is approximately Gaussian, according to the central limit theorem. The mean or expectation, E, of this distribution is the same as that of the input but the variance is that of the input divided by the number of terms in the filter, i.e.

$$E = \mu, \qquad v_n = \sigma''^2 = v/n \qquad \ldots(2)$$

where there are n terms in the filter. For example, a 64-term filter reduces the standard deviation of the input by a factor of 8, so giving a probability distribution like that of Fig. 27(b); the more terms in the filter, the more peaked the distribution. The extent of this distribution determines the steepness of the non-linearity, G, for, clearly, if G rises significantly over the range of the distribution then a substantial proportion of the picture will not be noise-reduced even when there is no motion.

It is interesting to compare the behaviour of the mean modulus approach with that of taking a root mean square. It is shown in Appendix 2 that the expectation and variance of the distribution formed by taking a finite r.m.s. value over $2n$ or $2n+1$ terms are given by

$$E_{2n}/\sigma = \frac{(2n)!}{(n!)^2}\frac{\sqrt{2n}}{2^{2n}}\sqrt{\frac{\pi}{2}}$$

$$E_{2n+1}/\sigma = \frac{(n!)^2}{(2n)!}\frac{2^{2n}}{\sqrt{2n+1}}\sqrt{\frac{2}{\pi}}$$

with $v_n = \sigma^2 - E_n^2$. These are plotted in Fig. 28.

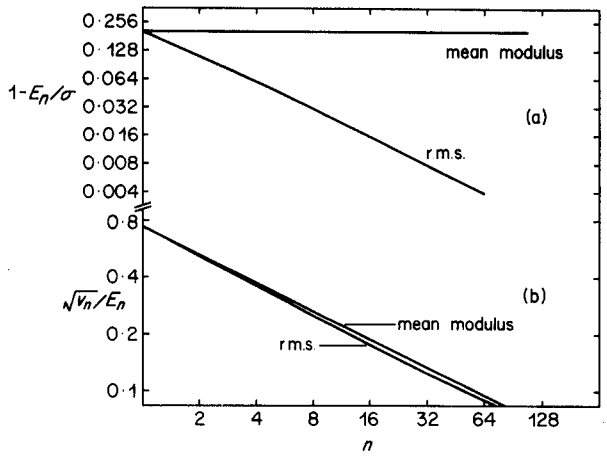


Fig. 28 – Statistics of r.m.s. and mean modulus values derived from n independent values taken from a normal distribution. (a) expectation (b) variance.

Thus as n increases the expectation of the finite r.m.s. value tends towards the true r.m.s. value and the variance diminishes. Moreover, as n increases, the ratio of the standard deviation to the expectation is less than that for the mean modulus case. This indicates that the r.m.s. approach is more efficient statistically. The reason why the r.m.s. approach was not pursued in practice is that it would have required double the number of bits in the digital processing.

3.3.3.3. The spatial filter

Given that the spatial filter forms an equal-weight sum of n terms there are two opposing factors governing the choice of n. The larger n becomes the more accurate is the measure of noise, as manifested by the lower variance of the output, and so the more readily can motion be detected by steepening the non-linearity G. On the other hand, the larger n becomes, the more slowly does the filter output respond to a change and therefore the more likely it is to miss isolated changes caused by small-area motion.

For example, consider a rectangular pulse

caused by a moving edge (ignoring camera integration) propagating through the filter. In the absence of noise the filter output would be a trapezoidal-shaped pulse as previously noted, being the convolution of the pulse and filter coefficient pattern. Because of noise, however, the output rises from a basal level, say E_0, with a statistical distribution about the trapezoidal shape. For a given pulse shape, the wider the filter, the earlier the output begins to rise but the less the difference between the peak output and the basal level, once the filter is wider than the pulse. On the other hand, the motion detector is more sensitive to this difference if the non-linearity is steeper. One would expect some compromise to emerge from these opposing factors and the behaviour can be analysed to some extent, as follows.

Assume that the filter forms an equal-weight sum of $2m + 1$ input samples. Then, letting the input and output be x and y, respectively, we have

$$y = (2m + 1)^{-1} \sum_{i=0}^{2m} x_i$$

Since the x_i have a statistically varying component due to the noise, the output, y, also has a statistical variation. The expectation of the distribution of y is given by

$$E(y) = (2m + 1)^{-1} \sum_{i=0}^{2m} E(x_i) \qquad \ldots(3)$$

where the $E(x_i)$ are the means of the distributions of the individual input samples.

In the absence of a movement signal, all the $E(x_i)$ are equal to, say, μ_0 where μ_0 equals $\sqrt{2/\pi}\sigma$ as previously noted.

Thus, for no movement

$$E(y) = E_0 = \mu_0 \qquad \ldots(4)$$

However a movement pulse of amplitude V will bias the probability distribution of an individual sample by the amount V as shown in Fig. 29. The mean of this distribution, as shown in Appendix 1, is given by

$$\mu = a\mu_0 + bV \qquad \ldots(5)$$

where $a = \exp - V^2/2\sigma^2$ and $b = \text{erf}\,(V/\sigma\sqrt{2})$.

The number of samples affected in this way will depend on the position of the movement pulse in the filter. If there are j samples of the pulse in the filter then the expectation of the filter output will be

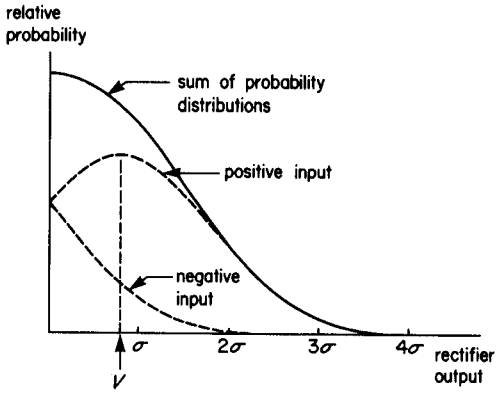


Fig. 29 – The probability distribution of the rectified picture difference signal assuming the difference is a constant, V, with added Gaussian noise of variance σ^2.

$$E(y) = (2m + 1)^{-1}\left[\sum_{i=0}^{j-1}(a\mu_0 + bV) + \sum_{i=j}^{2m}\mu_0\right],$$

using (3) and (5)

$$= (2m + 1)^{-1}[j(a\mu_0 + bV) + (2m + 1 - j)\mu_0]$$

$$= (2m + 1)^{-1}\{[2m + 1 - j(1 - a)]E_0 + jbV\},$$

using (4) ... (6)

Now if the pulse is $2p + 1$ samples wide the total filter response to the pulse is $2(m + p) + 1$ samples wide. Of this, only the central $2p + 1$ elements occur during the pulse (assuming the group delay is discounted) and thus the m elements either side give time for the motion detection to build up and die away. If the pulse is to be unaffected by the recursive noise reduction process then the spatial filter output at the $(m + 1)$th sample must be such as to have completely turned off the noise reduction. That is, the non-linearity, G, must have risen to unity by the $(m + 1)$th sample.

If the pulse width is greater than $m + 1$ samples then the expected value of the $(m + 1)$th sample of the filter output, s_{m+1}, is given by equation (6) with j equal to $m + 1$. Otherwise it is given with j equal to $2p + 1$. Thus there are two cases:

(1) $2p > m$,

$$s_{m+1} = [1 - (m + 1)(1 - a)/(2m + 1)]E_0 + (m + 1)bV/(2m + 1) \qquad \ldots(7a)$$

(2) $2p < m$,

$$s_{m+1} = [1 + (2p + 1)(1 - a)/(2m + 1)]E_0 + (2p + 1)bV/(2m + 1) \qquad \ldots(7b)$$

Now suppose we make the assumption that the non-linear characteristic, G, rises from its basal level, which prevails up to an input signal value of E_0, to unity at an input value which is three standard deviations (the classical 99.9% threshold) of the input distribution in the absence of motion, i.e. $3\sigma''$, above E_0. Fig. 30 shows the relationship between these functions. The standard deviation of the input, in the absence of motion, being that of the filter output, σ'', is given by

$$\sigma'' = \sigma'/\sqrt{2m+1} \qquad \text{using (2)}$$

Then by requiring the $(m+1)$th filter output sample to exceed the value which makes the non-linearity unity, i.e.

$$s_{m+1} \geqslant E_0 + 3\sigma'/\sqrt{2m+1}$$

and using equations (1) and (4) to express E_0 and σ' in terms of the r.m.s. noise, σ, we obtain an impairment threshold for the quantity V/σ, i.e. the amplitude of the movement pulse in units of r.m.s. noise, as a function of m and p, i.e. the filter and pulse widths.

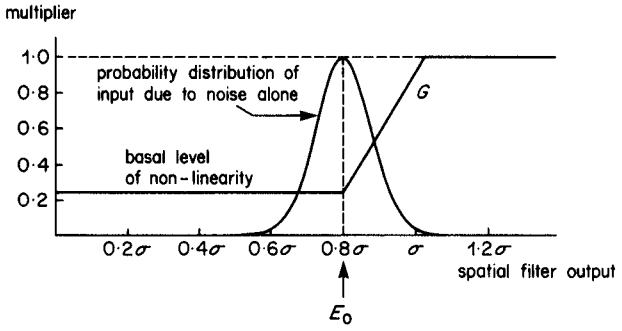


Fig. 30 – The relationship between the probability distribution of the spatial filter output and the non-linear function.

For $2p > m$, as can be seen from equation (7a), the value of s_{m+1} is independent of p. Letting V/σ equal z, it proves convenient, in practice, to express m in terms of z for this range whilst to express p in terms of z and m for the other range. Thus, after manipulation, we obtain:

For $2p > m$,

$$(m+1)^2/(2m+1) = c^2$$

with solution

$$m = c^2 - 1 + \sqrt{c^2(c^2-1)} \qquad \ldots(8a)$$

For $2p < m$,

$$(2p+1)^2/(2m+1) = c^2$$

with solution

$$p = \tfrac{1}{2}[c\sqrt{2m+1} - 1] \qquad \ldots(8b)$$

where

$$c = 3\sqrt{1-2/\pi}/[bz - \sqrt{2/\pi}(1-a)]$$

where

$$a = \exp -\tfrac{1}{2}z^2, \quad b = \operatorname{erf}(z/\sqrt{2}), \quad z = V/\sigma$$

The result is shown in Fig. 31 in which the minimum unimpaired pulse amplitude is plotted against pulse width, with filter size as parameter. For a given filter size, the amplitude of the pulse falls as its width increases until the width equals half the filter width. Thereafter, increasing the pulse width does not decrease the threshold amplitude. With large filters, small pulses are at a disadvantage compared with no filtering ($m = 0$), for which the threshold amplitude is 2.6 units. For example, pulses less than 4 samples wide ($p < 1.5$) are at a disadvantage with a filter 17 elements wide ($m = 8$).

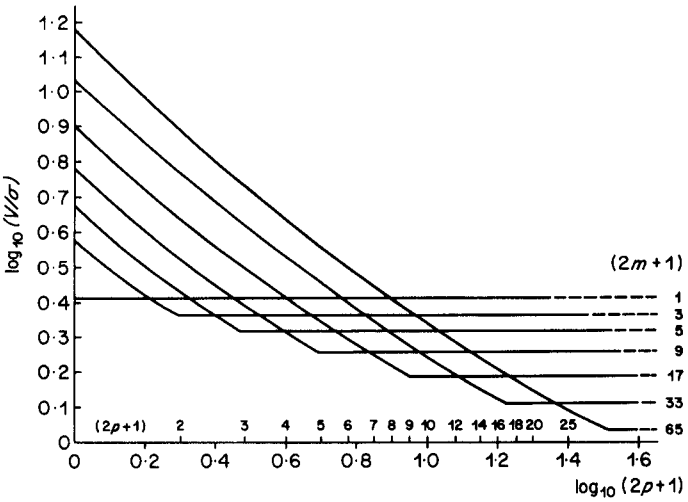


Fig. 31 – The relationship between minimum unimpaired pulse amplitude (in units of r.m.s. noise voltage) and pulse width for various widths of one-dimensional spatial filter.

Fig. 32 shows the effect of a two-dimensional filter of dimensions $(2m+1)$ elements in width by $(2n+1)$ elements in height. In this case the equations are just the same as before, assuming the pulse is one-dimensional, horizontally, except that

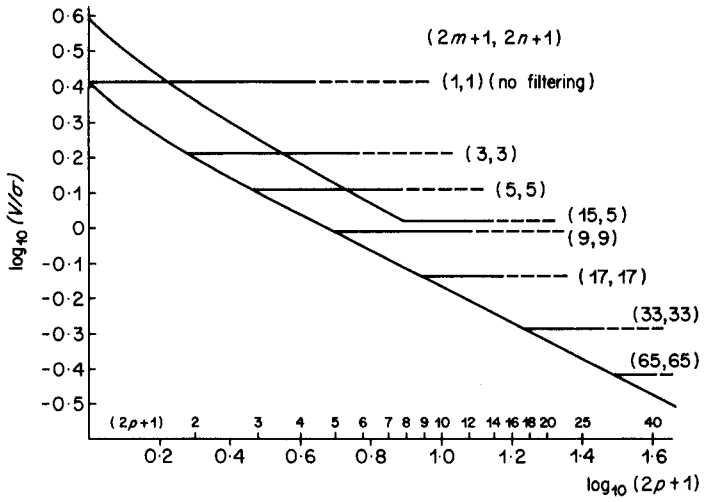


Fig. 32 – The relationship of Fig. 31 for various areas of two-dimensional spatial filter.

the standard deviation of the filter output, in the absence of signal, σ'', is reduced by a factor of $\sqrt{2n+1}$ and hence the steepness of the non-linearity is increased by the same factor. Thus the equations become:

For $2p > m$,

$$(m+1)^2/(2m+1) = c^2/(2n+1)$$

with solution

$$m = \{c^2 - (2n+1) + \sqrt{c^2[c^2-(2n+1)]}\}/(2n+1) \quad \ldots(9a)$$

and for $2p < m$,

$$(2p+1)^2/(2m+1) = c^2/(2n+1)$$

with solution

$$p = \tfrac{1}{2}[c\sqrt{(2m+1)/(2n+1)} - 1] \quad \ldots(9b)$$

As can be seen, the general effect is to depress the curves of Fig. 31 making the whole process more sensitive. But now, any square filter confers a detection advantage on any size of pulse compared with no filtering. However, shorter pulses are still harder to detect than longer ones, as might be expected. Moreover, filters whose widths are greater than their heights, e.g. the (15,5) filter in Fig. 32, are at a disadvantage compared with square filters whereas those filters with the reciprocal aspect ratio are better off. (This is a natural consequence of processing a horizontal pulse. The reverse would be true if the pulse were vertical.) Quite a large deviation from squareness, however, has little effect on the sensitivity for a given number of elements, once the pulse width exceeds the break point, as shown by the 81-element square filter (9,9) and the 75-element rectangular filter (15,5). As can be seen, the latter filter has a pulse width break point of 8 samples ($p = 3.5$) and a break point amplitude of about 1 unit (0 on the logarithmic scale). Below the break point, the threshold amplitude rises gradually to 4 units (0.6 on the logarithmic scale) as the pulse width decreases, only pulses of width one sample being at a disadvantage, compared with no filtering. The (15,5) filter was selected as it was convenient to instrument.

3.3.3.4. The non-linearity

The foregoing analysis is based on the assumption that the non-linearity rises from the basal level to unity over the $3\sigma''$ half-range of the input distribution, the "3σ" point being that below which 99.9% of the distribution lies. If the spatial filter were of infinite extent this distribution would be infinitely sharp and so the non-linearity would rise infinitely quickly. This section will show, however, that even if the filter were of infinite extent, there is a limit to the rate of change of the non-linearity. The limit occurs because the input to the spatial filter, in the absence of motion, is not just the rectified noise *input* to the recursive filter but the rectified sum of the input and *output* noise powers. The latter quantity depends on the amount of noise reduction which, in turn, depends on the value of the non-linearity and, hence, upon the input to the non-linearity. Thus we have a feedback loop in which the input to the spatial filter depends on its output, via the form of the non-linearity. This imposes a condition on the non-linearity which can be analysed as follows.

Consider the simplified circuit of the noise reducer shown in Fig. 33. Assuming for the moment that the spatial filter *is* of infinite extent, the input to the non-linearity, δ, has an infinitely thin probability distribution of expectation E_0, in

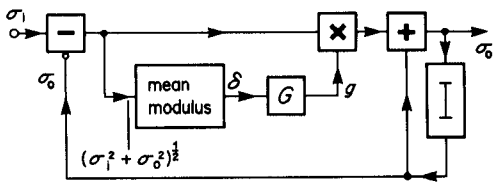


Fig. 33 – Simplified circuit of the noise reducer for analysing the dependence of its behaviour on the non-linearity.

the absence of motion, given by

$$\delta = E_0 = [(2/\pi)(\sigma_i^2 + \sigma_0^2)]^{1/2}$$

from equation (1) where σ_i and σ_0 are the r.m.s. input and output noise voltages.

But the output noise power is related to the input noise power through the noise factor, NF, which is a function of the recursive filter feedback factor, K, in Fig. 3. Here, the divider has been replaced by a multiplier, fed by the non-linearity, so that

$$K = g^{-1}$$

where g is the output of the non-linearity, G. Thus

$$(\sigma_0/\sigma_i)^2 = (2K - 1)^{-1}$$
$$= g(2 - g)^{-1} \quad \ldots(10)$$

whence $\delta = E_0 = (2/\pi)^{1/2}[2(2 - g)^{-1}]^{1/2}\sigma_i$

or $g = 2[1 - (2/\pi)(\sigma_i/\delta)^2] \quad \ldots(11)$

Thus we have derived a relationship between g and δ with σ_i as a parameter. But this relationship is already governed by the shape of the non-linearity, G. Hence we could solve for δ or g in terms of σ_i. However it is more instructive to keep the $g - \delta$ relationship explicit, to demonstrate what follows.

Fig. 34 shows a family of curves described by equation (11) with σ_i as parameter, together with a possible non-linear relationship between g and δ. For any given value of σ_i the values of g and δ are given by the point of intersection of the relevant σ_i curve with the non-linearity. Now it can be seen that, in general, a curve may intersect the non-linearity any number of times, depending on the nature of the non-linearity, so giving any number of stable states. A single state is obtained only if the slope of the non-constant part of the non-linearity is less than the σ_i curve at each point.

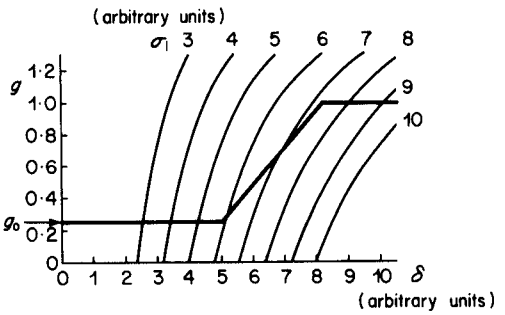


Fig. 34 – The dependence of the input to the non-linearity on its output as a function of noise input to the noise reducer together with a possible non-linearity characteristic.

For example, consider the discontinuous non-linearity of Fig. 35. Any input noise level corresponding to a curve within the shaded area gives three solutions for (δ, g), one corresponding to maximum noise reduction, one to partial noise reduction and one to no noise reduction. Input noise levels lower than this are maximally reduced, those above, unaffected.

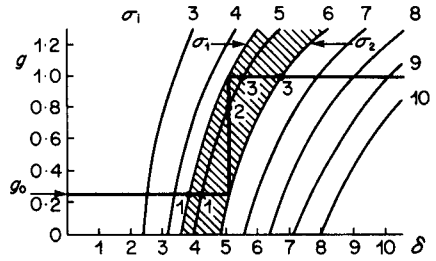


Fig. 35 – The same family of dependent curves as in Fig. 34 together with a discontinuous non-linearity characteristic.

This behaviour is one of hysteresis and can be appreciated by considering what happens to the input/output noise power characteristic. Fig. 36 shows this characteristic for the last-mentioned case. A hysteresis cycle might start by increasing the input noise from zero with the output noise following proportionally. At the input voltage σ_1 the lowest limiting curve in Fig. 35 is reached with the intersection point at the low level of the non-linearity. As the input noise is raised beyond this point the curves in the shaded area apply and the intersection point remains at the low level, noise reduction remaining effective until the input noise

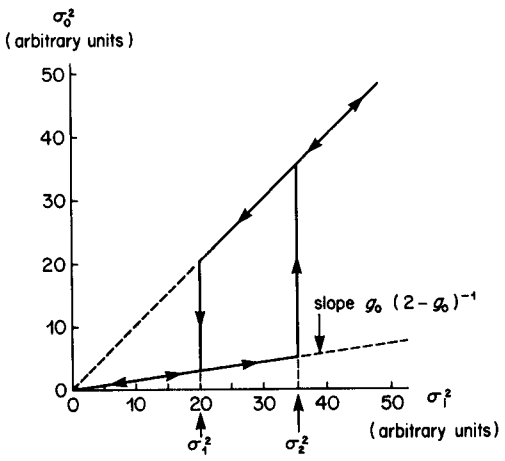


Fig. 36 – The dependence of the output noise power on the noise power input to the noise reducer, assuming the discontinuous non-linearity of Fig. 35, showing hysteresis.

reaches σ_2. Beyond this noise voltage the point must rise to the high level and noise reduction ceases so that σ_0 equals σ_i. Reducing the input noise now causes the intersection point to remain at the high level, with consequent absence of noise reduction, until the last possible moment when the input voltage σ_1 is reached, below which the intersection point drops to the low level and noise reduction becomes effective once more.

Clearly this hysteresis behaviour is undesirable and can be avoided by "backing off" the steepness of the non-linearity until it, at least, coincides with a σ_i curve. Coincidence with the curve for, say, σ_t, shown in Fig. 37, is the limiting case of stability and corresponds to a discontinuity in the input/output noise characteristic at σ_t as shown in Fig. 38. The arrangement is optimally efficient at detecting when σ_i exceeds σ_t or, provided σ_i remains at σ_t, at detecting motion.

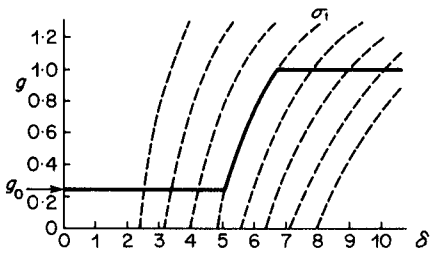


Fig. 37 – The same family of dependence curves as in Fig. 34 together with a limiting non-linearity characteristic.

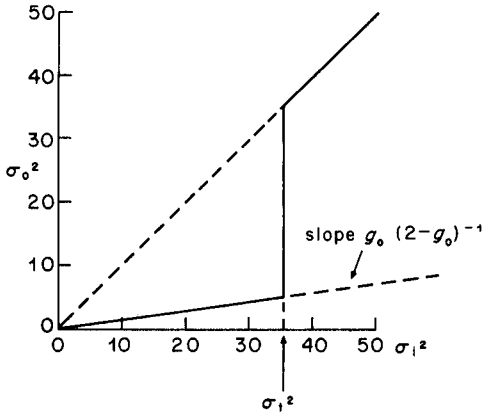


Fig. 38 – The input/output noise power characteristic of the noise reducer, assuming the limiting non-linearity characteristic.

Different values of σ_t can be accommodated by using the fixed relationship (equation (11) in another form)

$$g = 2[1 - 1/x^2] \qquad \ldots(12)$$

where $x = (\pi/2)^{1/2}(\delta/\sigma_t)$ and scaling the value of δ. Hence the need for the variable multiplier preceding the non-linearity. The universal plot of equation (12) in Fig. 39 shows that the ratio of the abscissae where the curve cuts the values unity and zero is $\sqrt{2}$, and any practical curve must have a ratio which exceeds this value.

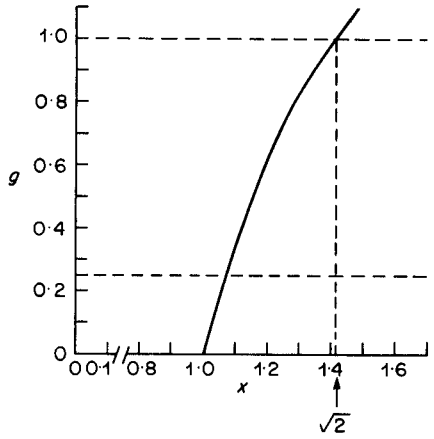


Fig. 39 – The limiting non-linearity characteristic expressed as a universal curve in terms of normalised variables.

In practice the limiting curve of equation (12) is not optimal subjectively, because the spatial filter is *not* infinite. This means that, for a constant input noise, δ is statistically distributed so that the working point on the curve is not constant but has a probabilistic behaviour, as previously noted. If the centre of the probability distribution lies at the lower breakpoint of the non-linearity where the curve cuts the line $g = g_0$, then the slightest deviation causes the instantaneous working point to extend up the curve and beyond it into the unity region. This appears as lumps of input noise in a background of reduced noise, whose size depends on the extent of the spatial filter.

This behaviour can be countered by lowering the working point so that the centre of the δ distribution lies below the lower breakpoint of the non-linearity, but this only decreases the probability of occurrence of the lumps, not their objectionability.

A better solution is to "back off" from the discontinuity in the input/output noise characteristic. Suppose a characteristic is defined which, for the sake of argument, has a linear transition of slope k between the noise-reduced and unprocessed regions, i.e.

$$\sigma_0^2 = g_0(2 - g_0)^{-1}\sigma_i^2$$

for $\sigma_i^2 < [1 - g_0 k^{-1}(2 - g_0)^{-1}]^{-1}\sigma_t^2$,

$$\sigma_0^2 = k(\sigma_i^2 - \sigma_t^2)$$

for $[1 - g_0 k^{-1}(2 - g_0)^{-1}]^{-1}\sigma_t^2 < \delta_i^2 < [1 - k^{-1}]^{-1}\sigma_t^2$,

and $$\sigma_0^2 = \sigma_i^2$$

for $[1 - k^{-1}]^{-1}\sigma_t^2 < \sigma_i^2$

as shown in Fig. 40. Then, using equations (11) and (12) to substitute for σ_0 and then σ_i we obtain, for the transition region,

$$g = 2k(1 + k)^{-1}[1 - 1/x^2]$$

where $x = (\pi/2)^{1/2}(\delta/\sigma_t)$ as before.

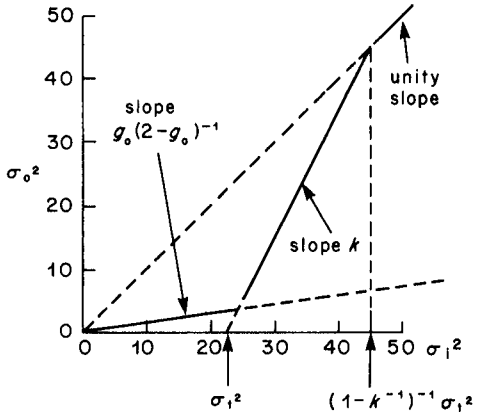


Fig. 40 – A possible input/output noise power characteristic for the noise reducer, having a linear transition region.

This relationship is plotted in Fig. 41. The curve for $k = \infty$ is the limiting case of stability corresponding to the discontinuous input/output noise characteristic of Fig. 38. The curve for $k = 1$ is asymptotic to unity as $x \rightarrow \infty$. For $k > 1$ the curve reaches unity when

$$x^2 = 2k/(k - 1) \qquad \ldots(13)$$

For example, the curve for $k = 2$ reaches unity when x equals 2.

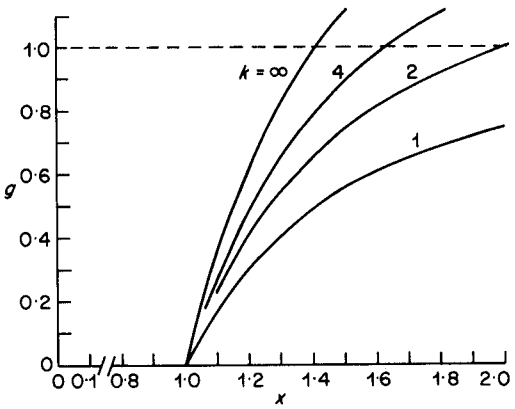


Fig. 41 – The normalised non-linearity characteristic resulting from the assumed input/output characteristic of Fig. 40.

The optimum value of k depends on the size of the spatial filter. We have shown that with an infinite spatial filter there is a limiting $g - \delta$ curve corresponding to $k = \infty$. On the basis of the previous theory, then, the appropriate value of k, with a finite spatial filter, is that which causes the non-linearity to rise to unity at an abscissal value of $3\sigma''$ beyond the limiting case ($\sqrt{2}$) where σ'' is the standard deviation of the filter output. For example, a 75-term filter gives an output with a ratio of standard deviation to mean of 0.086 or 8.6%. Thus if the mean is set at the lower breakpoint of the curve the "3σ" point is 26% beyond $\sqrt{2}$, i.e. at x equal to 1.78. The corresponding value of k is, from equation (13), 2.8.

In practice the optimum value of k was found by patient observation of pictures to arrive at an acceptable compromise between "hair-trigger" behaviour and smearing of low-level detail. The value finally arrived at was 1.6. However, it should be noted that the derivation of the non-linearity assumed a somewhat arbitrary linear relationship between input and output noise powers, in the transition region. Any other assumed relationship would have led to a different non-linearity which would have been equally valid. The chief value of the analysis in this sub-section (3.3.3.4) lies in the discovery of the limiting curve.

4. Colour operation

Given that the equipment must accept and deliver a composite colour signal there is a choice of two different system designs. One is to decode the incoming signal into its luminance and colour difference components, process the three signals by three nominally identical circuits like that of Fig. 5 and recode the three signals to PAL. The other is to use only one circuit with composite signals throughout but make due allowance for the picture-to-picture difference caused by the subcarrier. This difference is inherent in the definition of subcarrier frequency which is not an integral multiple of the picture frequency; if allowance were not made for this fact the offset would cause large difference signals in areas of high colour saturation. These would be interpreted as movement and thus saturated coloured areas would not be noise-reduced.

In such a method the composite signal in the picture store is transformed so that its colour information is appropriate to the input signal. This amounts, in the more complex case of PAL, to a 90° phase shift of the subcarrier and the reversal of the V axis switch, the operations taking place in a single 2-port network in series with the picture store, shown as a chrominance transformer in Fig. 42. In principle the operations could be carried out by decoding and recoding the signal at that point but the transformer would then need to contain fast digital multipliers for quadrature demodulation and remodulation together with the appropriate digital filtering and means for phase locking the multipliers to the burst. Whilst this would be the most general solution which could deal with non-mathematical composite signals, the obvious complexity can be avoided by approaching the problem in a different way.

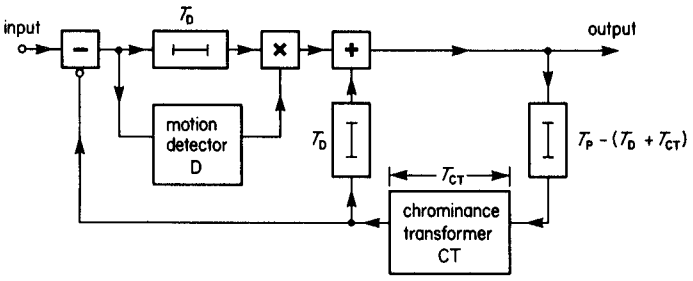


Fig. 42 – Simplified block diagram of the noise reducer showing the modification needed for composite colour signal processing.

Assuming the composite signal *is* mathematical, a different approach is to consider that the transformer is trying to predict the value of the current sample, based on samples occurring nominally one picture period earlier, i.e. the transformer becomes a predictor. This approach is that used in DPCM and, indeed, the noise reducer configuration of Fig. 42 is very similar to that of a DPCM coder. In these terms the monochrome noise reducer of Fig. 5 predicts the value of the current sample as that occurring precisely one picture period earlier, i.e. on the assumption that there is no movement in the scene at that point.

Turning to the composite signal case, however, this prediction can no longer be used because of the aforementioned offset in the colour subcarrier. Fig. 43 shows, in the PAL case, the phase of the subcarrier and V axis switch polarity relative to a central reference point at various sample positions in the previous field and picture. The sampling structure assumes an instantaneous sampling frequency of nominally three times the subcarrier frequency, but line-locked. As can be seen, the

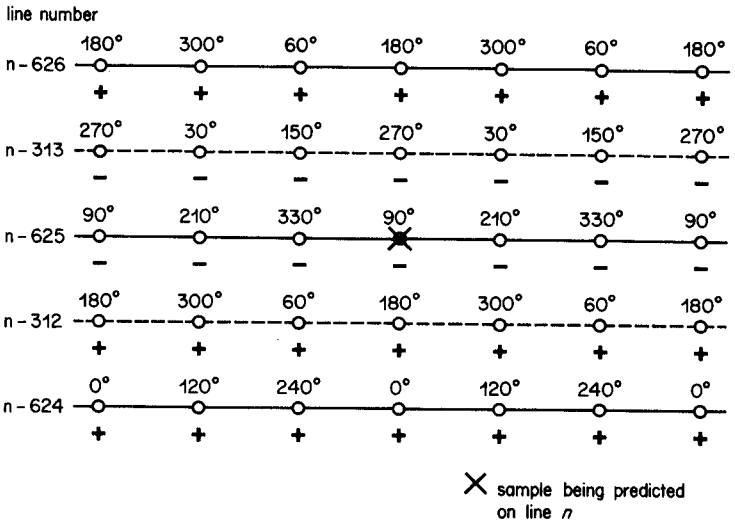


Fig. 43 – The phase of the PAL subcarrier and V axis switch polarity at sample points in the previous field and picture relative to a reference point in the centre.

phase of the subcarrier and the polarity of the V axis switch at the sample precisely one picture period earlier are both inappropriate. In fact the V axis switch reversal precludes the use of any samples from the line containing that sample. On the other hand, the sample exactly 624 line periods earlier has the appropriate subcarrier phase and V axis switch polarity. (Actually the phase is not precisely the same but is in error by 0.576° as a consequence of the picture frequency term. This is too small to be shown in Fig. 43.) Thus this sample can be taken as a prediction for chrominance components, provided that the positional error can be tolerated. This sample also gives a good prediction for line-locked luminance frequency components in the chrominance band which give rise to cross colour of fine vertical frequency. These observations are independent of the actual subcarrier phase and V axis switch polarity at the predicted sample which, thus, need not be known. Moreover, as samples either side of the predicted sample are not involved, the precise value of the sampling frequency is irrelevant, the only condition being that the samples are in vertical register from picture to picture.

If the chrominance band is defined by a single bandpass filter, then using the z-transform notation the required transfer function of the predictor, H_p, is

$$H_p(z) = (1 - B)z^{-625} + Bz^{-624} \quad \ldots(14)$$

where z^{-1} is the transfer function of a line delay and B is the transfer function of the bandpass filter. The first term in equation (14) represents the

luminance part of the prediction for low frequencies and the second term represents the chrominance part for high frequencies, the exact form of B to be determined by experiment. This transfer function can be realised by the circuit of Fig. 44 in conjunction with the picture store as in Fig. 42, which must now correspond to a delay of nominally 624 line periods, less the group delay of the bandpass filter. (In fact the store must also be shortened by the group delay of the movement detector as shown in Fig. 42, but this is a separate issue.)

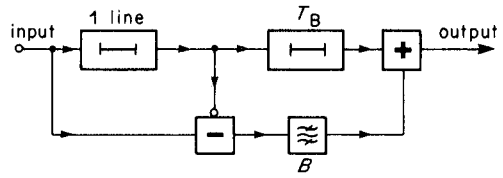


Fig. 44 – The circuit of the PAL predictor taking chrominance contributions from only one line.

Such a predictor fails where there is a sudden change of colour from one line to the next because of the positional error. To mitigate this effect the sample above the prediction point can also be taken into account to provide a mean. As shown in Fig. 43 this has the opposite subcarrier phase and so must be inverted in taking the mean. Although this gives a better chrominance prediction the high frequency luminance prediction is worse because of the inverted contribution. Specifically, the prediction for line-locked luminance is now zero. Whether or not this predictor is preferable to the other depends, amongst other things, on the relative weight of the chrominance and high-frequency luminance contributions, in average picture material.

Using the same notation as before, the transfer function of this predictor is given by

$$H_p(z) = (1 - B)z^{-625} + B\tfrac{1}{2}(z^{-624} - z^{-626})$$

which can be realised by the circuit of Fig. 45.

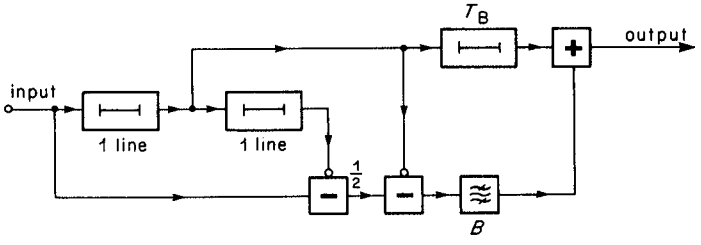


Fig. 45 – The circuit of the PAL predictor taking chrominance contributions from two lines.

Both these predictors were realised experimentally and used in the configuration of Fig. 42 to process composite signals in real time. The most serious drawback of the first predictor then became apparent. This was a systematic upwards movement, at the interlace strobe speed, of some of the residual noise left after the noise reduction process. This gave an impression of rising steam and was quite objectionable even though caused by only that part of the noise in the chrominance band defined by the bandpass filter. Thus the second predictor, which gave no such effect, was chosen for subsequent development. Varying the spectral characteristic of the bandpass filter, B, was found to have little effect on the efficacy of the noise reduction process, subject, of course, to the filter's having unity gain at subcarrier frequency. If anything, the movement impairment tended to be less when the characteristic was narrow. Ultimately it was decided to make the characteristic approximate to that of the luminance notch in the PAL decoder. This meant, in practice, making the narrowest characteristic possible without significant spectral overshoot with the amount of hardware available. Fig. 46 shows the characteristic obtained.

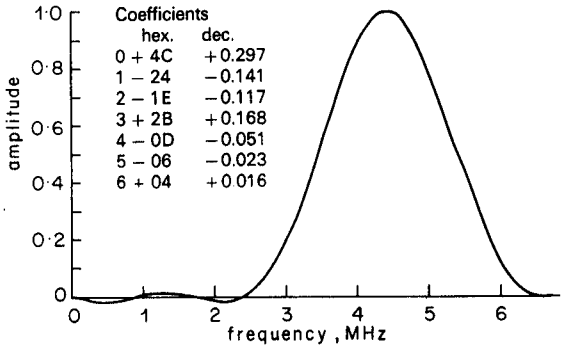


Fig. 46 – Frequency characteristic of the bandpass filter used in the PAL predictor.

Turning to the NTSC case, the arguments are similar except that there is no V axis switch and the subcarrier phase shift across one picture period is 180°. As this is also the phase shift across one line period it follows that the sample in the previous picture either above or below the predicted sample position has a suitable subcarrier phase. As far as the chrominance component is concerned, therefore, the mean of these samples is preferable to avoid the "rising steam" effect and to provide some interpolation for line-to-line chrominance changes. Moreover, as both contributions are positive the mean also provides a better approximation to the high-frequency luminance component, if line-locked, than in the PAL case.

The transfer function of the predictor in this case, assuming a 525-line scanning system, is

$$H_p(z) = (1 - B)z^{-525} + B\tfrac{1}{2}(z^{-524} + z^{-526})$$

which can be realised by the circuit of Fig. 47 in conjunction with a nominal picture delay. The form of B was, again, kept as narrow as possible and Fig. 48 shows the characteristic used.

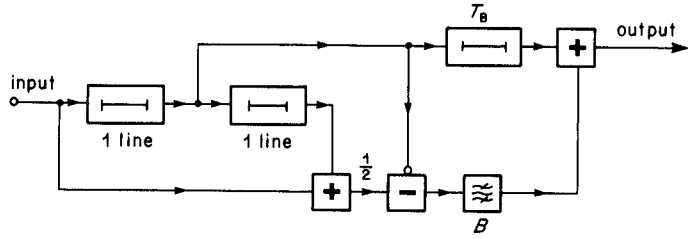


Fig. 47 – The circuit of the NTSC predictor.

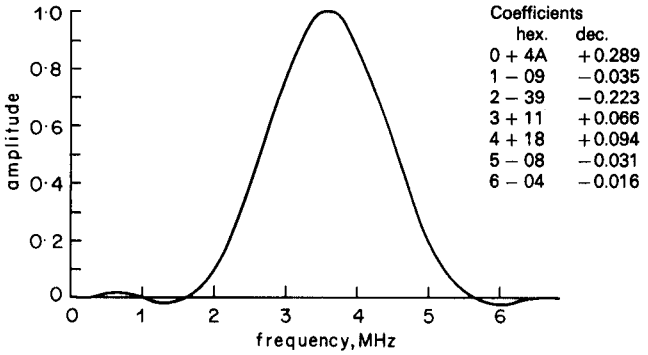


Fig. 48 – Frequency characteristic of the bandpass filter used in the NTSC predictor.

5. The practical realisation

This Section discusses some of the factors affecting the design of an equipment based on the foregoing principles.

5.1. Picture store

At the outset, the picture store represented the overriding cost of the hardware. The experimental prototype equipment used a store based on 1 kbit shift registers[4] which was readily available and had been used for previous investigations. This was constructed as two one-field stores, each contained in a six-foot high bay, and capable of working at clock rates up to 13.5 MHz. Unfortunately the integrity of the store was rather dubious in that the automatic test routine rarely indicated the complete absence of store errors. Fortunately, in the noise reducer, store imperfections are mistaken for movement and so are partially "protected" by the movement detector, resulting, merely, in failure to noise reduce, the more catastrophic the fault, the greater the likelihood of protection. This, in fact, proved to be the case, and frequently the equipment could be operated for quite long periods with quite serious store errors before their effects were noticed.

5.2. Clock generation

The need for an exact picture delay in the recursive filter, at least for the low frequency luminance, implies the use of a digital clock frequency which is picture frequency locked so that it gives picture-stable samples. Moreover the theory of the colour predictor implies the use of a line-locked frequency. A frequency of 851 times the line frequency was used as clock generators working on this frequency were readily available. Such generators use the line synchronising pulses to steer an oscillator in frequency and phase and care must be taken to ensure that the stability of the pulses is adequate. Correct operation of the predictor requires that the clock frequency is stable to an accuracy of a few degrees of subcarrier phase, and this can only be achieved by using very long time constants in the oscillator control loop to average out the jitter in the separated pulses. In addition the oscillator must have a high degree of basic stability. This part of the equipment is crucial and, during the course of the project, no less than four clock pulse generators were designed. The first generator was based on an easily-available LC oscillator and was quickly found not to be stable enough. Subsequent designs were based on crystal oscillators with means for digitally rephasing the count of the line-frequency divider, to deal with non-synchronous cuts.

5.3. Circuit block

Fig. 49 shows a block diagram of the actual circuit of the noise reducer so far described. All the processing elements were based on TTL except for the line stores which were based on 1 kbit MOS shift registers. These had the capacity to store only the active line plus part of the back porch and so the whole recursive filter had to be bypassed by a path taking the unprocessed part of each line.

The spatial filter in the motion detector, described in section 3.3.3.3, was constructed as cascaded vertical and horizontal filters with provision for manually varying the number of terms in each up to the maximum of 5 and 15 respectively. The vertical part was constructed in straightforward transversal form so that the line delays in the filter, being substantial, could also act as the

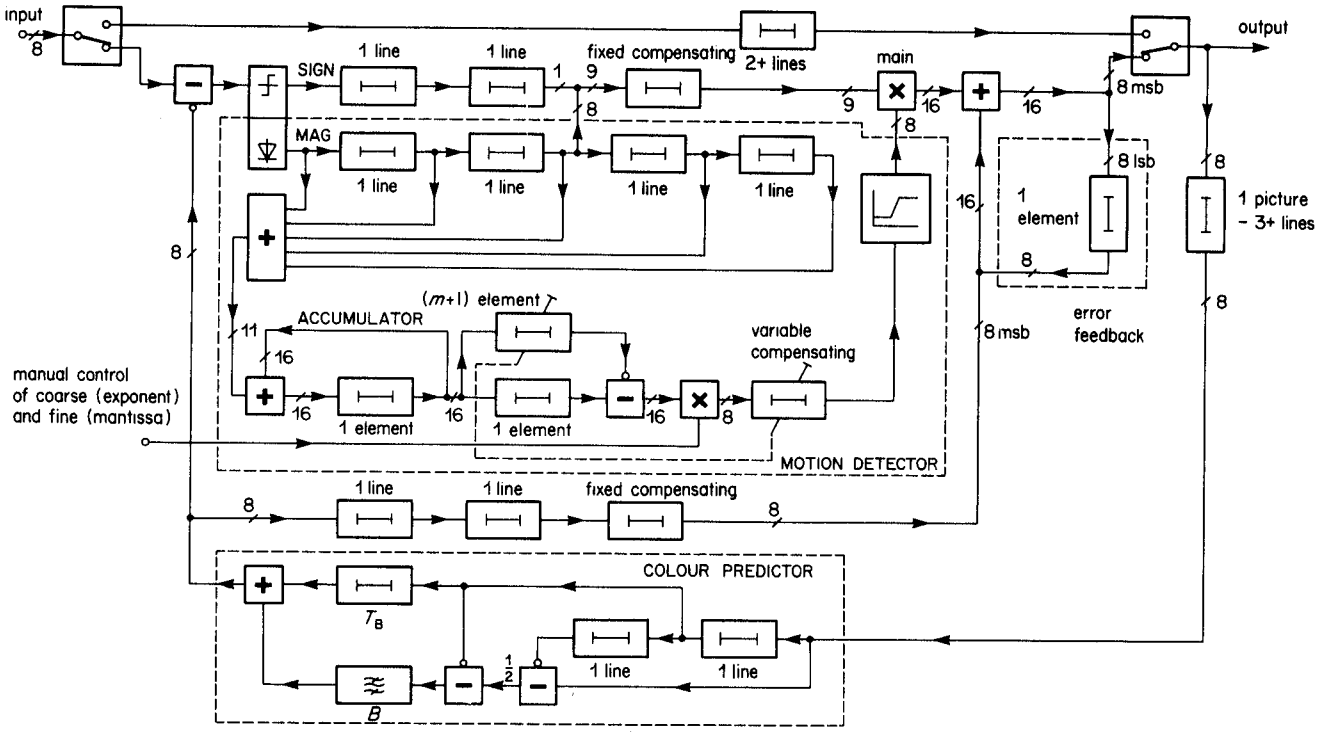


Fig. 49 – Block diagram of the complete noise reducer.

compensating delay for the main path signal. The rectification of the picture difference signal was performed by selecting the magnitude output of a sign/magnitude converter, the sign being retained for delay compensation but not passing through the ensuing filter. Selection of one, three or five terms was performed in the summing network.

The horizontal part of the spatial filter was realised using an accumulator or integrator, followed by subtraction across a delay corresponding to a number of horizontal picture elements. This technique gives a running sum of samples over an aperture defined by the delay even though the accumulator continually overflows, provided modulo arithmetic is used. The variation in group delay as the filter delay varied was compensated by a following delay, varying in a complementary way.

As the whole spatial filter effectively adds together up to 75 terms, extra data bits are generated. The vertical filter summing network output contains three extra bits and the accumulator output, a further five bits. This poses a problem for the multiplier, following the spatial filter, which must therefore operate on a 16-bit signal. The solution was to include a limiter before the multiplier and to realise the multiplier in floating point form with independent control of binary exponent and mantissa as shown in Fig. 50. The exponent control acts in conjunction with the

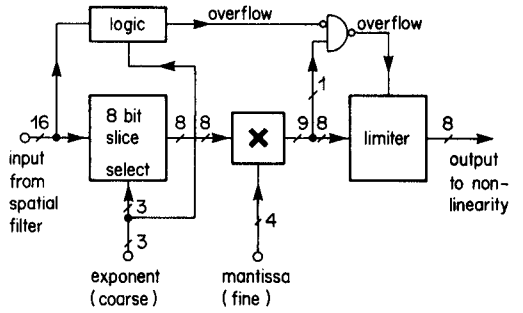


Fig. 50 – Block diagram of the motion detector multiplier.

limiter by selecting a "window" of 8 bits to be passed on to the mantissa multiplier and, at the same time, testing the input to see whether or not it will be over-range after shifting. The three-bit exponent allows multiplication in the range 2^{-3} to 2^{4}. A second over-range test is applied after multiplication by the four-bit mantissa which lies in the range one to two. If the result is over-range through either of the mechanisms then it is limited to the maximum value which can be expressed using the eight-bit output. The exponent and mantissa controls were brought out to remote binary switches to form a coarse and fine control of the motion detector.

The non-linearity was contained in a programmable read-only memory (PROM) having a 256 × 8 word format, so arranged that the break point

of the curve lay at the input value of nominally 64. (Actually this is the value where the curve cuts the horizontal axis.) In the absence of motion this is the value that should emerge from the multiplier/limiter when it is correctly adjusted. For research purposes alternative non-linearities were provided in various PROMS, selectable by a manual control.

Finally, care had to be taken to ensure that the group delay of the filter, multiplier and non-linearity path was equal to that of the main path so that the output of the non-linearity entered the main multiplier at the time appropriate to the difference signal arriving at the multiplier via the main path. This was ensured by adjusting the fixed delay in the main path to the main multiplier. Once this delay was fixed, it set the length of the fixed delay in the predictor feed to the main adder so that both signals reached the adder at the appropriate times. This delay, in turn, set the length of the nominal picture delay as the loop length round the picture store, predictor and adder must be one picture period.

5.4. Mechanical details

The equipment was built using standard BBC Binary Metric Module size 4U PCB's, each 156 mm high, for the circuit just described together with BBC-designed analogue to digital and digital to analogue converters. In addition, a manual bypass was provided, operated by a front panel switch. The board packing density was highly variable as befitted an experimental prototype and no attempt was made to compress the design to a more compact form. The assembly of racks fitted into a 6 foot high, 19 inch wide cabinet, slightly smaller than each of the two field store cabinets. Fig. 51 shows an overall view of the equipment; all the circuitry except the field store is contained in the right hand of the three cabinets, i.e. in the centre of the picture, with the analogue/digital conversion at the bottom and the power supplies at the top.

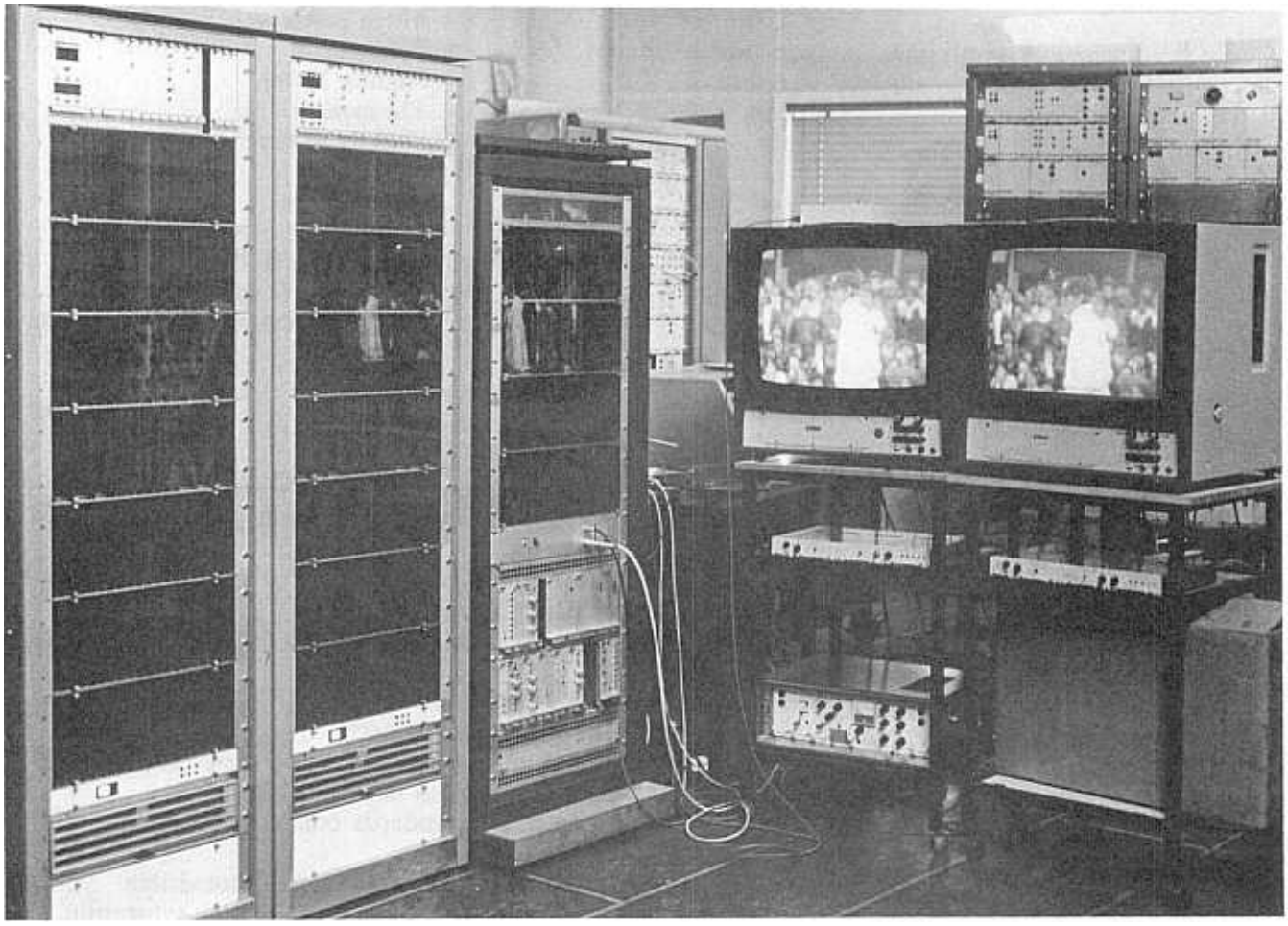


Fig. 51 – Overall view of the experimental noise reducer.

6. The first field trial of the equipment

Once the equipment was built it was imperative to test its performance "in the field" on as wide a variety of programme material as possible. Such a test was also intended to bring to light difficulties in installation, operation and maintenance. Accordingly the equipment was installed in Television Centre in the Standards Conversion Area for about a fortnight in June 1977 with the intention of using it on tape-to-tape transfers during the post-production phase of programme generation. In the event it was used mainly live in the network output feeds after an initial trial period. In addition, a test was carried out after normal broadcasting hours in which selected viewers were asked to judge the processed pictures received in their homes.

The specific points on which guidance was needed were on (1) "dirty window", the almost stationary pattern formed by residual noise, (2) adequacy of movement portrayal, particularly in texture, (3) the possibility of finding a compromise setting for the motion detector control for a given noise level to allow "hands off" operation, (4) the types of programme material most likely to benefit from noise reduction, (5) the required response of the equipment to non-standard waveforms and non-synchronised cuts between video signals.

The "out of hours" test transmitted on June 22nd, 1977 consisted of nine excerpts shown in Table 2 which represented different kinds of noisy source material except for two which were relatively quiet.

Viewers were asked to judge:

1. The effectiveness of noise reduction using the CCIR Rec. 500 quality scale:

Grade	
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

2. Loss of sharpness in moving areas using the CCIR Rec. 500 impairment scale:

Grade	
5	Imperceptible
4	Perceptible but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

3. Overall quality of the processed picture compared to that of the unprocessed picture using the CCIR Rec. 500 comparative scale.

Grade	
+3	Much better
+2	Better
+1	Slightly better
0	The same
−1	Slightly worse
−2	Worse
−3	Much worse

The results of this test compiled from 84 completed forms, of which 47 were from viewers to Crystal Palace and the remainder from widely

1. Royal Variety Performance	An outside broadcast with an excessive amount of low-frequency noise.
2. Table Tennis	An outside broadcast with highly saturated blue areas displaying a high level of "chrominance" noise.
3. Sailor	16 mm film shot under available light. Some scenes probably 2 stops under-exposed.
4. Black and White Minstrels	Well lit with very little source noise visible.
5. Survivors	High noise level from lightweight cameras.
6. Lion Tamer	Well exposed 16 mm film representative of best 16 mm film performance. Grain perceptible but not disturbing.
7. Money Programme in America	Good original 525-line V.T., standards converted to 625 at TV Centre. Usual added converter noise.
8. Ballet Sequence	Outside broadcast from Covent Garden. Noise quite visible.
9. Morecambe and Wise	Multi generation V.T. Moiré patterns in areas of high saturation.

Table 2 : Picture material used for "out-of-hours" test

spread transmitters, are shown in Fig. 52 on the left side.

It will be seen that the noise reduction was generally rated as average to good except for items 4 and 6 which were the two relatively quiet excerpts. Loss of sharpness was hardly noticed even for those items requiring heavy noise reduction. The overall rating was again very encouraging except for the quiet items.

In an attempt to discover whether or not the signal distribution network contributed significantly to the noise the results from 13 viewers at the ends of rebroadcast links were analysed separately and the results of this are shown on the right hand side of Fig. 52. It can be seen that these results are broadly similar to the others with perhaps a slight dilution of the variation. This showed that the noise reduction was still worthwhile even though some viewers commented that they were in a poor signal area.

In addition to the out-of-hours test a log was kept by Research Department staff to answer the questions concerning the performance of the equipment. The following points were noted as a result of the records, observations by Television Centre staff and the results of the out-of-hours test:

1. "Dirty window" was sometimes noticed when the input noise was greater than −32 dB unweighted (ref. 0.7 V).

2. Movement portrayal of fine texture, even when optimised, was sometimes not acceptable. Reappearance of noise in moving areas was, however, never noticed.

3. A compromise setting for the motion detector was impossible to achieve because of variation of the incoming noise level, not only from programme to programme but from shot to shot. In studio material, camera-to-camera variation was at least 3 dB and, likewise, with film, the varying conditions of exposure caused variations in the level of film grain. Multi-generation video tape noise and standards converter noise did not, however, vary so much and so were much easier to deal with.

4. All types of programmes seemed to benefit from noise reduction, particularly those derived from electronic cameras working under low light conditions.

5. The performance of the equipment on non-sync cuts was not really tested due to its being on the network output for most of the time. On non-standard waveforms it merely failed to reduce noise. Residual picture store faults also caused failure to reduce noise in moving bands which was sometimes noticeable.

In addition to the above points two further points should be noted. These are:

6. Noise level varies significantly over the grey scale with studio cameras and, to a lesser extent, with film. The variation is not predictable as it depends on the setting of the gamma correctors and aperture correctors in camera channels as well as the inherent noise properties of the source. This variation gave rise to differential noise reduction wherein some parts of the picture could be noise-reduced and others not.

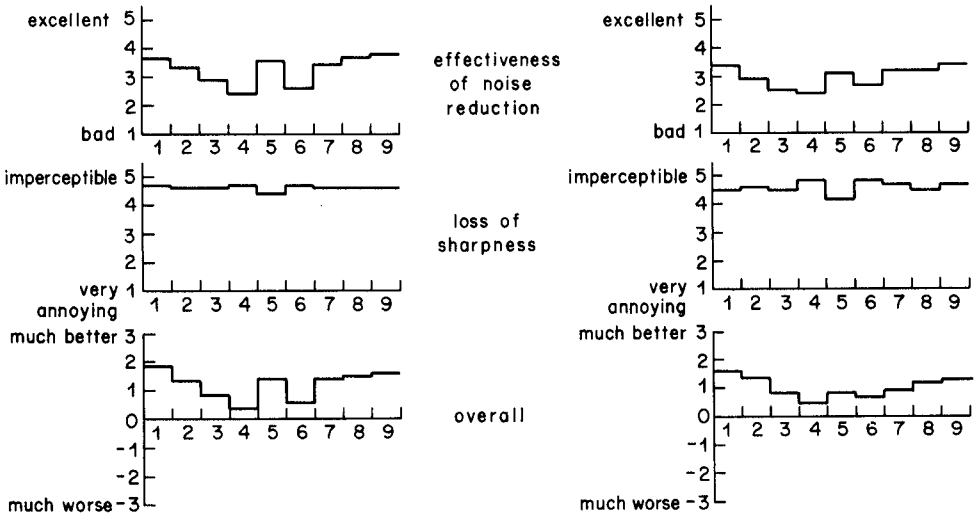


Fig. 52 – The results of the "out-of-hours" test.

7. Motion of low-level detail was sometimes not detected, although motion of adjacent higher level detail was detected. This gave rise to dissociation of the picture in local areas which was particularly disturbing when caused by unsteadiness, for example the slow "bounce" of 35 mm film.

As a result of these observations, the following conclusions were reached concerning required improvements:

1. Account had to be taken of the varying distribution of noise over the grey scale.

2. Automatic setting of the motion detector was essential under operational conditions.

3. Movement detection had to be improved to avoid dissociation. Automatic operation would help this to some extent.

4. Operation on the 525/60 standard was desirable to cope with noise introduced by standards converters working in the 625–525 direction.

5. A more reliable and, preferably, compact picture store was desirable.

In general the reliability of the equipment was very encouraging and the fact that it was used "live" on programmes for long periods enabled it to be used and evaluated effectively. It caused little critical comment while its noise reduction performance was judged to be very satisfactory.

7. Automatic adaptive operation

7.1. Introduction

The first two requirements of the preceding list were met by developing an automatic system for adjusting the motion detector control. This control affects the value of the multiplier between the spatial filter and the non-linearity in Fig. 26 (conveniently referred to as the side-chain multiplier) whose function is to keep the signal input to the non-linearity constant at the lower break-point in the absence of motion. When there is no motion the spatial filter output is proportional to the noise voltage and so the required multiplier value is inversely proportional to the noise voltage.

Assuming the noise voltage can be measured, there are two ways of effecting the control of the multiplier based on either feed-forward or feedback principles. In the feed-forward mode, the measurement is made on the spatial filter output. This is then applied to a reciprocal circuit which produces a multiplier value for the side chain multiplier. This method is stable but has the disadvantage that the controlling circuit has to handle 16 bit signals from the spatial filter which involves undue instrumental complexity.

In the feedback method the measurement is made on the side chain multiplier output. An error signal is formed from the measurement during one field and used to generate a correction signal which, in turn, generates the multiplier value for the next field. This method avoids the instrumental complexity of 16 bit signals because the limiter in the side chain multiplier ensures that its output contains only 8 bits. However, in common with all closed loop systems, its stability needs to be considered.

Either method needs a reliable way of measuring noise in the presence of motion and detail and this turns out to be the core of the problem. In practice it was found to be extremely difficult to do and it governed the effectiveness of the whole machine.

Variation of noise over the grey scale is dealt with by replicating the control circuit and selecting the appropriate circuit according to the instantaneous luminance value of the signal being noise-reduced. The number of circuits required depends on the rate of change of noise level with grey level, together with the accuracy expected of the side-chain multiplier value. Four was chosen as this was thought to be the minimum number needed to characterise the different forms of variation that would be met in practice.

Each control circuit thus handles the derivation of the multiplier value for a range of luminance and the range boundaries must be chosen with due regard to the expected variation. Inevitably the choice will be a compromise. In practice it was found that on most signals the noise tended to vary more at black than at white and so the range boundaries were closer at the black end of the luminance scale.

Fig. 53 shows how the feedback control system is connected to the rest of the noise reducer. A feed of the side-chain multiplier output is taken to a range switch which distributes the signal to the appropriate control circuit and the appropriate output is selected and fed to the multiplier. Control of the distribution and selection is derived from a luminance level selector which divides the luminance into the predetermined ranges. Luminance is derived by a simple chrominance-rejecting filter

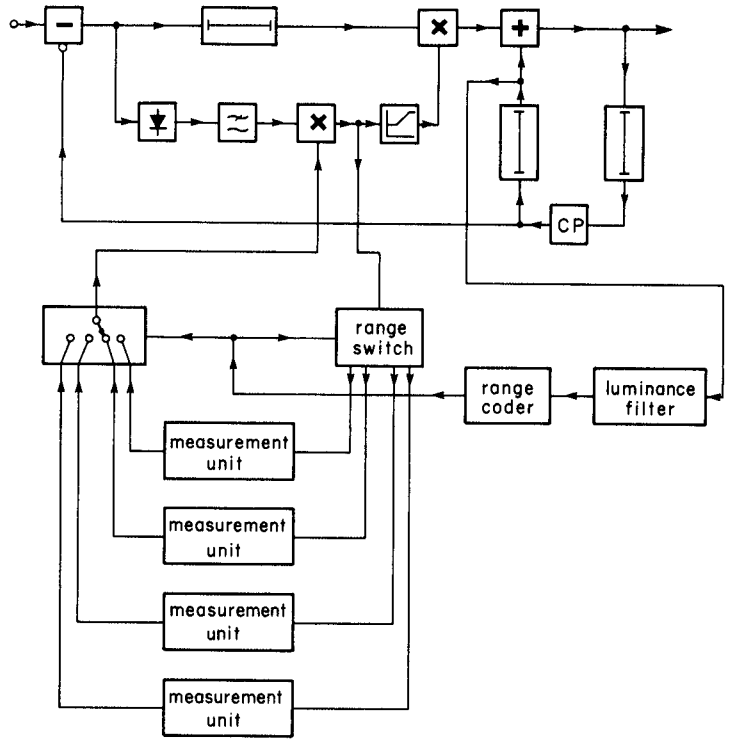


Fig. 53 – Block diagram of the motion detector control system and its connection to the noise reducer.

taking a feed of the processed output signal, appropriately timed.

7.2. The feedback control circuit

In practice, the feedback control method was chosen to avoid handling signals with too many bits. As mentioned earlier, the measurement of the noise output of the side-chain multiplier on one field is used to derive the multiplier value on the next field. The multiplier is of floating point form, as explained in Section 5.3, so that it can handle a large dynamic range. However, the theory of the multiplier control loop assumes that the multiplier is of logarithmic form so as to ease the realisation. This assumption modifies the transient response of the loop but does not affect the quiescent state.

A block diagram of the control loop is shown in Fig. 54 and its action can be explained as follows. Suppose the quiescent noise signal on the preceding field is n and, on a particular field under consideration, becomes $n + \Delta n$. Then the multiplier value held in the store for the preceding field is assumed to be $\log_2 (n_0/n)$ where n_0 is the required clamping level for the noise i.e. the input level at the lower break point in the motion detector non-linearity. The output of the multiplier, corresponding to the noise level, is then n_0/n times $n + \Delta n$ which is $n_0(1 + \Delta n/n)$ and this is the value registered by the noise measurement circuit if it is operating correctly. From this, n_0 is subtracted to leave $n_0 \Delta n/n$. This error signal is then applied to a read-only memory to produce the value $-\log_2 (1 + \Delta n/n)$ which is the amount by which the stored multiplier value must be incremented to give the required multiplier value of $\log_2 [n_0/(n + \Delta n)]$ for the field under consideration.

The mathematical law stored in the PROM is ideally of the form $-\log_2 (1 + x/x_0)$, shown in Fig. 55 but in practice it is modified as shown. This is

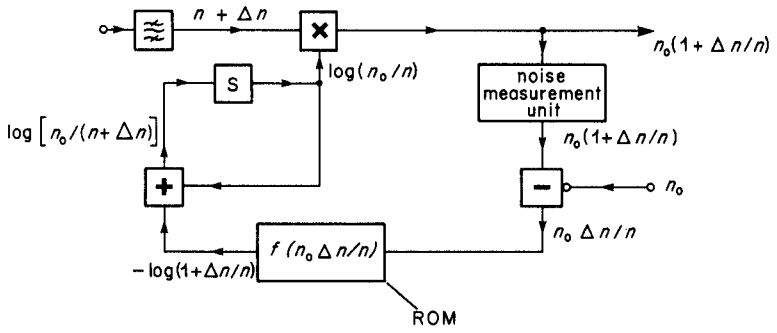


Fig. 54 – Simplified block diagram of the motion detector multiplier control loop.

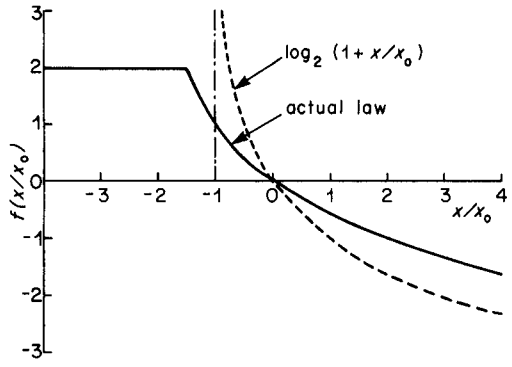


Fig. 55 – The ideal and practical forms of the non-linearity contained in the PROM in the control loop.

because, firstly, a limit is applied as $\Delta n/n$ tends to -1 since the logarithm tends to infinity. This limit has the effect of limiting the rate at which the control system responds to a drop in noise level beyond about 18 dB. In practice this is not noticed because the eye takes time to get accustomed to a new shot with a change in noise level. In the same way, the maximum input to the PROM corresponds to an increase in noise level of about 12 dB beyond which the slew rate of the control loop is again limited, and not noticed for the same reason.

Secondly, a scale factor is applied to lower the loop gain because, for increases in noise level, the required incremental correction is actually smaller than it would appear. This is because the transient increase in noise is interpreted as motion before the system responds and so the noise reduction action stops, thereby making Δn appear larger than it really is.

These factors, in combination with the approximation of the floating point form to the logarithmic form, mean that the mathematical law stored in the PROM is somewhat of a compromise. However, they only affect the transient response as previously noted and, in practice, the system appears to work well enough.

7.3. Noise measurement and derivation of multiplier values

7.3.1. The basic idea

The method of noise measurement is based on the assumption that the lowest value of the spatial filter output represents the noise level. This follows because the output is a smoothed, rectified picture difference which contains components due to both noise and motion but which is always unipolar. In practice this lowest value is a statistical quantity because it is derived field-by-field from a quantity which has a probability distribution function as described in section 3.3.3. Fig. 56 shows a hypothetical function which might occur with and without motion. From this it can be seen that the minimum value of the spatial filter output should give an indication of noise level which is fairly independent of motion but its precise relationship to the mean of the distribution, in the absence of motion, is unknown.

In theory the minimum value of the spatial filter output could be zero but the probability of this happening is negligible since this corresponds to about 11 standard deviations below the mean in the absence of motion if the input samples to the filter are all uncorrelated. The actual minimum obtained, in the absence of motion, will obviously depend on how many samples are taken at a time, the more taken, the lower the value obtained. Fig. 57 shows the probability distribution function of the minimum of n uncorrelated samples taken from a normal distribution for a wide range of n. (The derivation of these curves is given in Appendix 3.) This indicates that a minimum taken over eight samples has a mean value of about 1.4 standard deviations below the mean of the sampled distribution whereas one taken over 1000 has a mean of about 3.1. Thus the number of uncorrelated samples contributing to a minimum has a marked effect on the value obtained. Moreover, these values have their own distributions, becoming

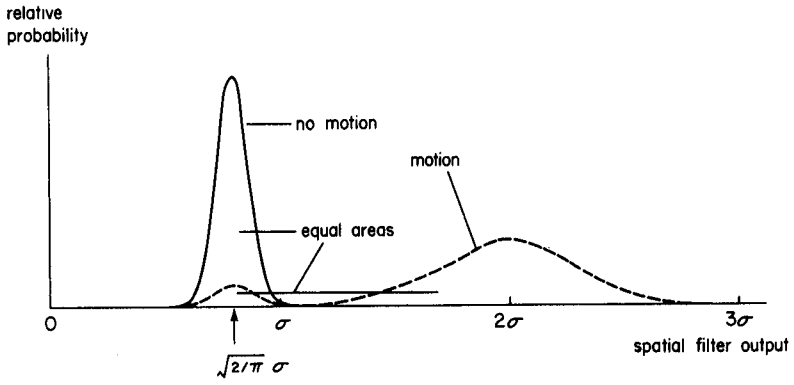


Fig. 56 – Examples of the probability distribution function of the spatial filter output.

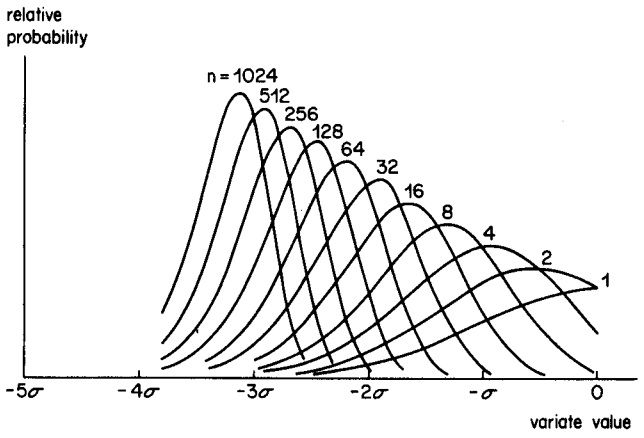


Fig. 57 – The probability distribution of the minimum of n uncorrelated samples taken from a normal distribution with zero mean and variance σ^2.

sharper as n increases. Thus a minimum obtained from 1000 samples is less variable than one obtained from eight. However, the decrease in variance is not proportional. This implies that for a given number of samples it is better to divide them into groups and form several minima and average them than to form one minimum. The number of samples contributing to any one minimum must be high enough, however, to ensure a sensible value since the method relies on the assumption that in the presence of motion there are still parts of the picture where the picture difference is caused solely by noise. These two factors, namely the number of samples contributing to a minimum and the number of minima averaged are thus in conflict, and a compromise must be sought.

In practice the compromise is achieved by minimisation throughout a television line and averaging 256 such line values to give a measurement once per field. This assumes that somewhere in each line there is a region of no motion. Although the number of samples in each line is about 690 the line-minimum value so derived does not correspond to this value of n since the smoothing of the filter imposes a high degree of correlation on adjacent output samples. An estimate of the value of n may be derived by assuming that the output samples in a block, equal to the width of the filter, are totally correlated. Then, as the filter width is 15 samples, the value of n is about 46. According to Fig. 57, this gives a mean value of about 2 standard deviations below the mean of the filter output, in the absence of motion, with a standard deviation of about 0.4. Averaging these values over 256 lines reduces the standard deviation by a factor of 16 to about 0.025. Thus the result of the combined minimisation and averaging operations is to produce a value of about 2 ± 0.025 standard deviations below the mean.

The above analysis ignores the fact that successive output samples of the filter are only partially correlated so that the figure of 46 is an underestimate. Moreover, the five-sample vertical extent of the filter introduces some correlation between successive line minima so that the variance reduction factor of 256 is an over-estimate. At one extreme every filter output sample could be considered uncorrelated with its neighbours both horizontally and vertically; at the other extreme the minimisation and averaging could be considered to take place only over samples separated by the filter dimensions. The first extreme leads to the value 3 ± 0.019, the second, to the value 2 ± 0.056. Clearly the dominant uncertainty is caused by the unknown horizontal correlation but, for any given situation, the value obtained will be very consistent.

A drawback of the method is that the measurement is in terms of the standard deviation of the original probability distribution function. This, in turn, depends on the spectrum of the noise since the averaging effect of the spatial filter depends on the degree of correlation between the input samples. The smallest standard deviation occurs when the input samples are all uncorrelated i.e. when the noise is white. As the correlation increases the smoothing action decreases and the standard deviation increases. Thus the measured noise value decreases as the noise spectrum departs from white. This adds a further degree of uncertainty to the measurement in any given situation.

7.3.2. Protection against motion

A noise measurement circuit based on the minimisation and averaging principle was constructed and tested. For this purpose the side chain multiplier value was monitored using a digital indicator. Motion was simulated by shortening the length of the nominal picture delay in the main recursive filter. The BBC Test Card F was used as picture material, and blanked noise was added to the signal before entering the filter.

With no simulated motion the measurement circuit appeared to work well, giving a stable measurement. However, with added motion the measurement increased consistently at all levels of noise. This indicated that the underlying assumption on which the measurement was based is not true – there is no region in each line where the picture difference is unaffected by motion. Clearly this result will depend on picture content for, in the limit, any featureless parts of the scene will be unaffected. Since the Test Card has such areas it

must be concluded that they are not big enough to allow the spatial filter output to settle to a quiescent state for long enough.

This fundamental difficulty is overcome by making the assumption that in any sequence of pictures the motion will, sooner or later, stop. Thus, by taking the minimum field-by-field measurement, i.e. by temporal minimisation, the true noise value will be obtained. This is conveniently done by allowing the control loop to respond only to decrements in the measured noise level. It should be noted that the noise measurement now consists of four processes: pixel-to-pixel averaging (spatial filtering), horizontal minimisation, line-to-line averaging, temporal minimisation. This last process is termed "ratcheting".

The ratchet action must not, however, override legitimate increases in measurement brought about by increases in noise level. These are most likely to occur at a shot change when the pictures are derived from a new source. So the ratchet action must be inhibited after a shot change for long enough to allow the new measurement to be established. If the change in noise level is substantial then the establishment of a new measurement may take several fields due to the limited slew rate of the control loop, mentioned in section 7.2. In practice the worst-case response time was found to be about eight fields so, for safety, the ratchet action is enabled after 16 fields have elapsed from the shot change.

Detection of shot changes is based on a refinement of a method described by Sanders and Weston.[5] This looks for characteristic behaviour of the field integral of the rectified field-to-field difference. To form this quantity the output of the spatial filter is integrated, for convenience, since the smoothing of the filter has little effect on the integral. Fig. 58 shows a simplified block diagram of the shot change detector. The picture-difference field integral is clocked into an eight-stage shift register at field rate so that eight successive values are simultaneously available. The average of the central pair of values, scaled and offset by a fixed amount, is then compared with the average of the six neighbouring values. In the absence of motion an ideal shot change produces two successive large values in the midst of small neighbouring values which would be easy to detect. However, in practice, motion and noise produce a varying sequence of samples so that the comparison of the two averages is less clear cut. Moreover, imperfect film splices introduce a further degree of difficulty. The scaling and offset of the central values give protection against false detection caused by these effects. A refinement is the subtraction from the input of a quantity proportional to the noise voltage. This is derived in a pan detector circuit to be described.

The strategy of allowing upwards adaptation to a higher noise level for only 16 fields after a shot change works well for the majority of the time but it is defeated when the shot change is to a scene with much motion such as a pan. Under these circumstances too high a level of noise would be registered and consequently the motion would be smeared. It is necessary, therefore, to detect such a situation of global motion and inhibit the upwards adaptation until the motion ceases. More particularly, any field after a shot change during which global motion is below a threshold is allowed to contribute towards the 16 during which upwards adaptation is allowed.

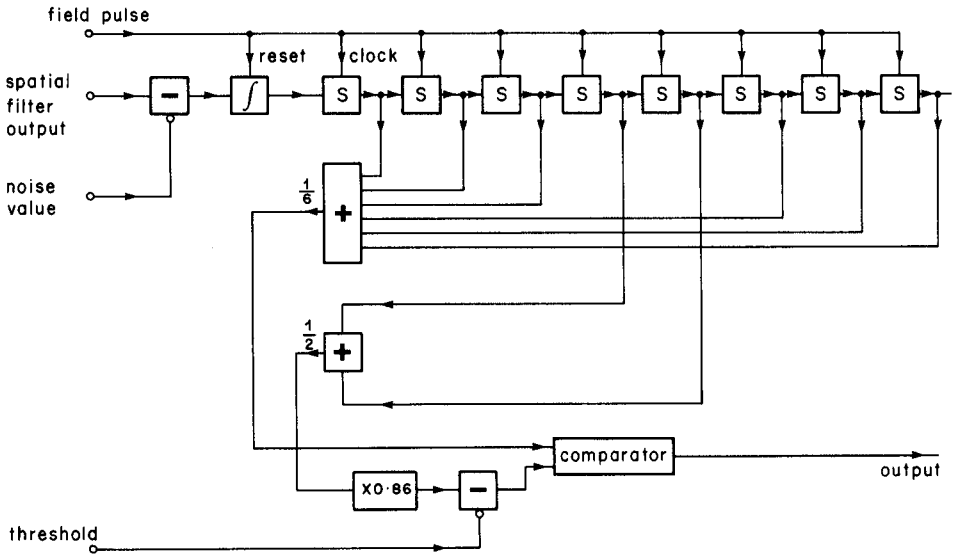


Fig. 58 – Simplified block diagram of the shot change detector.

Detection of global motion was initially based on the smoothed value of the field-integral of the rectified picture-to-picture difference. This is the predicted value derived in the shot change detector from the six neighbouring fields. However, after a second field trial of the equipment, this method was superseded by an improved method based on an idea by Storey and Roberts.[6] This uses the observation that the rectified picture difference, integrated along the line, varies from line to line more with global motion than without. Thus the difference between the maximum and minimum integrals registered during a field constitutes a measure of global motion.

Fig. 59 shows a simplified block diagram of the global motion detector. The rectified picture difference is integrated and reset once per line, the input being, once again, the spatial filter output since the filter smoothing has little effect on the line integral. The sampled and held (boxcar) line-integral signal is then subjected to maximisation and minimisation which are reset once per field. The difference between the running maximum and minimum is then formed and compared with a threshold to form a binary decision about global motion. In addition, the output of the minimum detector is sampled at the end of each field and used as a rough value of noise measurement, suitably scaled, to offset the input signal to the shot change detector mentioned earlier.

The effectiveness of the global motion detector clearly depends on the value of the threshold with which the difference between maximum and minimum is compared. The limit of sensitivity is set by the highest noise level expected and by the average error in stationary pictures caused by the colour predictor circuit.

A fundamental problem remains with a shot change to a pan with a lower noise level because even if a lower noise level is registered it will still be too high. So the downwards adaptation will be insufficient and movement smearing will result.

7.3.3. Grey scale dependence

As indicated in Section 7.1, the noise measurement and multiplier derivation is replicated four times to cope with variation of noise over the grey scale. This introduces further complications into the measurement process.

Firstly, each measurement circuit cannot simply be enabled whenever the luminance falls into the appropriate grey range. This is because the spatial filter output takes time to settle to a new value of noise (in the absence of motion) due to the averaging effect. Fig. 60 shows an idealised version of what happens at a luminance boundary, assuming that the noise decreases with increasing luminance. The noisy input signal is shown at (a) and the

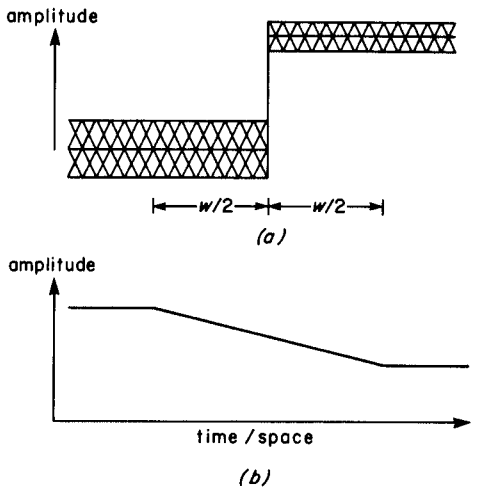


Fig. 60 – The effect of noise variation at a luminance boundary (a) noisy input signal (b) spatial filter output.

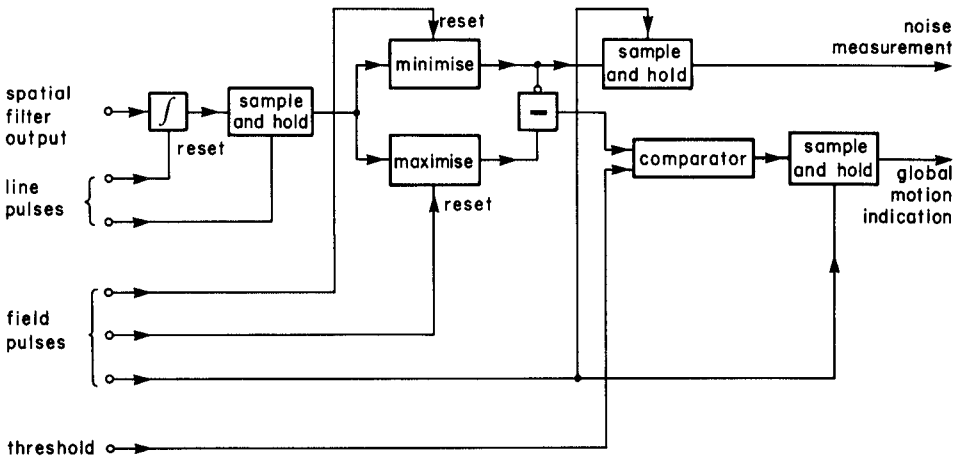


Fig. 59 – Simplified block diagram of the global motion detector.

spatial filter output at (b). It can be seen that the output of the spatial filter begins to change at a point which is half the width, w, of the filter before the boundary and only settles at the new value at the corresponding point after the boundary. The minimiser would thus measure a lower value of noise than the true value in the area of higher noise level. Moreover if the excursion into the area of lower noise were short enough, too high a level of noise would be recorded there. This would not be corrected until a large enough area occurred on the same line. In addition, because the spatial filter has a vertical extent the same effects occur if the luminance changes from one line to the next.

To guard against these false measurements the measurement process must be inhibited near the range boundaries. These boundaries, together with the edge of the picture, divide up the picture into areas. If a guard space, equal to half the filter dimensions, is allocated on both sides of each boundary, the areas, over which measurement is allowed, shrink. This is termed "windowing" and is the function contained in the range switch of Fig. 53. It is conveniently instrumented by passing each range signal through four line delays to give simultaneous access to five vertically adjacent signals and forming a five-input AND function so creating the vertical part of the window. The horizontal part is then formed by setting a flip-flop only after 15 consecutive indications occur, as determined by a counter, but resetting it whenever an indication disappears.

The grey level distribution of natural scenes, together with windowing, reduce the number of measurements that fall in any particular grey range during a line. The number may be reduced to a handful or, in the limit, to none at all. As Fig. 57 shows, the noise measurement is heavily dependent on the number of contributions to the minimum and so a lower limit to this number must be applied to prevent the minimum being too high. If the limit is not reached the noise measurement made on that line is declared non-bona-fide.

In the same way the distribution of grey levels may be such that the number of bona-fide line measurements within a field is reduced to too small a number for confidence to be placed in the averaged minima. Again, a limit is applied below which the field is declared non-bona-fide.

This complication of non-bona-fide lines and fields for each grey range modifies the noise measurement process. Fig. 61 shows a block diagram of the complete noise measurement circuit for each grey range. The input from the sidechain multiplier is now minimised only when the "measurement enable" signal, derived by the windowing technique, is valid. At the same time the "measurement enable" signal feeds a resettable pixel counter which raises a flag signifying that the line is bona-fide if the signal is valid for a predetermined number of contiguous pixels. If the line is bona-fide the minimiser output is taken for all valid pixels along the line; if not, the value n_0 is substituted. The line minimum is fed to an accumulator which effectively averages the line values until 256 lines have elapsed. At the same time the bona-fide flag signal feeds a line counter which keeps a record of the number of bona-fide lines. If this number, L, exceeds a predetermined value the counter raises a flag signifying that the field is bona-fide.

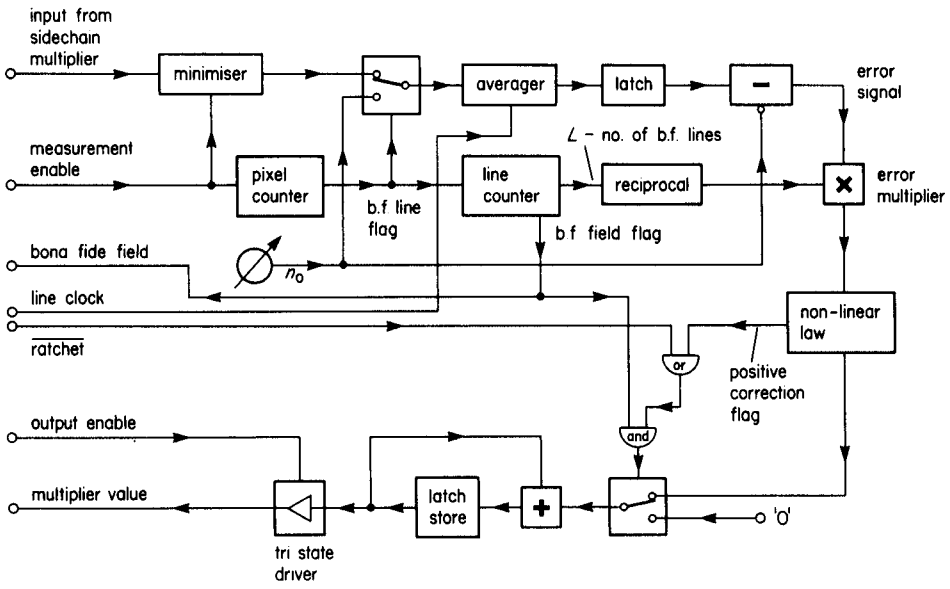


Fig. 61 – Block diagram of the noise measurement unit.

At the end of the field the value n_0 is subtracted from the averaged line values to form an error signal. If the error contains any contributions from non-bona-fide lines, however, its value will be smaller than if no such lines had occurred by a factor of $256/L$. This, in turn, would increase the time constant of the feedback control loop. To mitigate this effect the error signal is multiplied by a factor depending on the number of bona-fide lines. Ideally this factor is $256/L$ but, to avoid a complex digital multiplier, the factor is restricted to powers of two which is easy to implement by bit shifting. To avoid oscillatory control loop behaviour, caused by a loop gain of greater than unity, the factor is chosen to be the power of two below the value of $256/L$ unless L, itself, is a power of two.

The increment to the multiplier value resulting from the error is conditionally applied to one input of an adder, the other input of which receives the previous side-chain multiplier value held in the latch store. The output of the adder, representing the next multiplier value, is clocked into the latch store at the appropriate time, ready for the next field. The addition of the increment is conditional upon the correct combination of the error sign, bona-fide field and ratchet signals. The logic gates function so as to allow incrementing only when the field measurement is bona-fide and negative incrementing only when the $\overline{\text{RATCHET}}$ signal is valid. This last signal is generated from the shot change and pan signals, as previously described, except that the fields counted after a shot change must not only have sub-threshold global motion but must also be bona-fide. Thus a feed of the bona-fide field indication must be taken to the ratchet generating circuits and the $\overline{\text{RATCHET}}$ signals for the four grey ranges will, in general, differ.

Finally, the multiplier value so obtained and available in the latch store throughout the next field is selected by the output enable signal which is valid whenever the luminance of the processed signal falls in the grey range. Selection is by means of a tri-state device which connects the latch output to a bus feeding the side-chain multiplier.

7.3.4. Additional safeguards

The strategy of allowing upwards adaptation to a higher level of noise only after shot changes, conditional upon sub-threshold global motion and bona-fide measurements obviously fails if the shot changes cannot be detected. This occurs when the changes are brought about by "fades" or "wipes", for the "signatures" of these events are difficult to detect.

This problem is overcome, to a large extent, by generating artificial shot changes at regular intervals of five seconds if a real shot change does not occur for 10 seconds. A subsequent real shot change stops the generator and re-initiates the measurement of shot-to-shot interval. The artificial shot changes are protected by the global motion criterion in the same way as a real shot change, thereby preventing momentary smearing that would result if upwards adaptation occurred in a fast moving scene.

If the noise measurement for any grey range is not bona-fide on a particular field then it is ignored and the sidechain multiplier is not altered, as explained above. At the same time the multiplier value of the nearest bona-fide range is substituted on the multiplier bus whenever the luminance falls in the non-bona-fide grey range. This substitution continues, even though the range may become bona-fide in successive fields, until the next shot change, either real or artificial.

Occasionally, three or more grey ranges becone non-bona-fide. This often occurs at programme junctions where shots are faded to and from a single luminance level. In such circumstances the adaptive system is reset by detecting such a condition and inserting an artificial shot change. As the shot change is protected by the global motion criterion, noise reduction resumes when the fade has finished.

8. The re-engineered equipment

The prototype equipment used in the first field trial, described in Section 6, was completely re-engineered in preparation for a second field trial. The re-design took into account those factors brought to light by the first field trial listed at the end of Section 6. The most significant developments were the inclusion of an automatic system for measuring the noise and adjusting the motion detector, as described in Section 7, and a considerably smaller picture store. These developments enabled the whole equipment to be contained in one bay.

8.1. The picture store*

As a result of developments in integrated circuit technology a charge-coupled device capable of storing 64 K bit became available. As a complete PAL television field contains about 2.13 Mbit at the chosen sampling frequency it can be seen that only 32 devices will store the active portion and most of

* The picture store was designed and commissioned by A. Oliphant.

the rest, provided they are fast enough. These can be assembled on a single 4U board and such an economy of space was thought to be well worth achieving. In fact the development was somewhat speculative as, at the time, the devices were not being manufactured in volume and those used were pre-production samples. However, in other respects they appeared to be ideally suited to the application since their shift register organisation, with low power consumption, was exactly what was needed.

The maximum clocking frequency of the device was quoted as four MHz. This implied a four-way demultiplex at the chosen sampling frequency, giving 32 parallel bit streams, one device per stream. As a result, serial-to-parallel and parallel-to-serial converters operating on the input and output data, respectively, were required.

The charge-coupled device, itself, is organised as a serial-parallel-serial structure with 16 parallel registers, each of capacity four K bits. Data is stored by filling each register successively, using a four-bit address to determine which register is used at any instant. Data is shifted in all 16 registers simultaneously on receipt of clocks and reading takes place at the same location as writing. The complex clocking requirement of the total arrangement means that the 32 storage devices carry an overhead in clock-driving devices of 12 together with 15 further devices associated with addressing and input and output multiplexing. Nevertheless, it still proved feasible to mount all the components on a single 4 U board with a power consumption of about 14 W. A general view of the board is shown in Fig. 62.

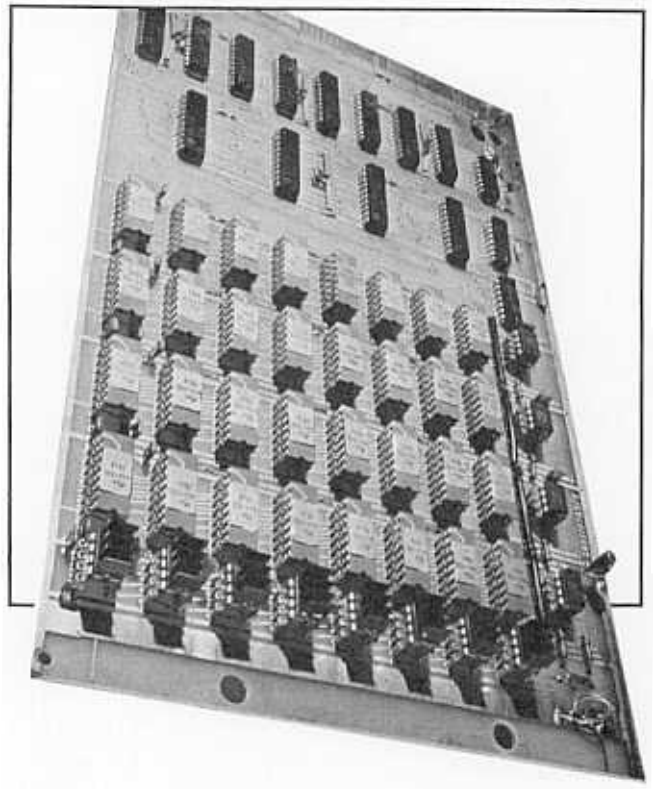


Fig. 62 – General view of the compact field store board based on 64 K CCDs.

The clock waveforms and addresses for both field store boards are generated on a separate housekeeping board. Trimming of the delay offered by the stores is accomplished by omitting clock pulses during the line blanking interval and, for NTSC operation, which requires significantly less storage than PAL, omitting CCD addresses. These operations are again performed on the housekeeping board.

In operation the store was found to be fairly reliable but several CCDs failed in the course of two or three years and had to be replaced. Eventually, it became impossible to obtain further devices and failed devices were allocated to least significant bits of the store where their isolated errors would cause least damage.

8.2. The clock generator*

As a result of the experience gained on the first field trial the clock generator was completely redesigned and particular attention was paid to the behaviour in the event of non-sync cuts. The part of the generator dealing with this aspect was further modified after the second field trial, to be described. The generator is based on the concept of a crystal oscillator locked to the incoming line syncs using a charge pump controlled by a line sync phase comparator. The need to deal with both 625/50 and 525/60 standards requires two entirely separate generators since the change in line frequency, and therefore of sampling frequency, being 851 times this in both cases, cannot be accommodated using a single crystal.

The clock, in addition to feeding the rest of the noise reducer, is used to generate housekeeping waveforms at line rate by clocking a counter, the output of which addresses a set of PROMS, containing the waveforms. One of these waveforms, a line pulse, is used to clock a further set of counters once per line. The outputs of these counters are used to address two pairs of PROMS, one pair containing picture rate housekeeping waveforms for the 625/50 CCIR Television System I, used in the UK, and the other pair, waveforms

* The clock generator was designed and commissioned by P. Fraser.

for the 525/60 System M, used in the USA and Japan.

In the event of a non-sync cut the line and field rate counters are re-phased so causing minimal disturbance to the phase comparator, and allowing a quick response time. In the first design a non-sync cut was detected by the appearance of a sync pulse outside a time window surrounding the expected position. This approach proved, in practice, to be unreliable as it could be affected by spurious sync edges caused by impulsive interference. In the second design the identification of a non-sync cut was made to take more time and be more reliable thereby.

A fundamental design conflict arises between the length of time taken to identify the cut and the shift of oscillator frequency brought about by the continual phase error during the identification period. In the second generator the time constant of the phase locked loop is made long enough for the phase error to have a very small effect on the frequency over the identification period. A non-sync cut is defined as a situation where 256 sync pulses fall outside the time window within a certain period. This period may vary from 256 line periods to about four seconds under extreme conditions.

Detection is achieved by allowing the first misplaced sync pulse to trigger a monostable which enables a counter to keep a total of the number of subsequent misplaced pulses during the monostable period. The monostable period is somewhat greater than 256 line periods but it can be re-triggered so that the counter remains enabled provided that the pulse separation does not exceed the monostable period. This caters for the situation where the sync misplacing is marginal so that not every pulse is misplaced. With a well-defined non-sync cut the counter reaches its terminal count within one field. An isolated burst of interference, on the other hand, is unlikely to satisfy the detection conditions since it tends to create widely separated groups of a few pulses.

When the cut is detected a reset pulse is generated which re-phases the line rate housekeeping counters and also triggers a second monostable. This monostable prevents any further re-phasing and pulse counting for two seconds, long enough for the clock oscillator to establish phase and frequency lock.

8.3. The predictor circuit

The requirement to deal with 525/60 NTSC as well as 625/50 PAL implies, as well as a change of housekeeping waveforms, a switched colour predictor circuit. Figs 45 and 47 show the circuits required for both cases and, from these, it can be seen that the only structural difference is between an adder and a subtractor. However, the bandpass filter characteristic must also change in sympathy with the subcarrier frequency according to Figs 46 and 48. This implies that the coefficients of the transversal filter must be remotely switchable.

The change from adder to subtractor is easily accomplished by selecting inverted or non-inverted data. The change of filter characteristic is achieved by passing coefficient values continually to the coefficient stores on the transversal filter boards, the coefficient set being stored on a separate filter housekeeping board. Different sets may then be selected according to operational requirements.

8.4. Miscellaneous mechanical details

Fig. 63 shows the single-bay layout of the re-engineered equipment. As can be seen, the digital processing, with the exception of the analogue/digital conversion, is confined to two racks, each of 20 boards. The upper rack is concerned with the "main path" through the noise reducer including the movement detector and picture store whilst the lower rack is concerned with the colour predictor and the automatic noise measurement and control system. This latter system occupies about half the rack, comprising four noise measurement boards, one for each grey range, three control boards and two line-delay boards needed for measurement window generation.

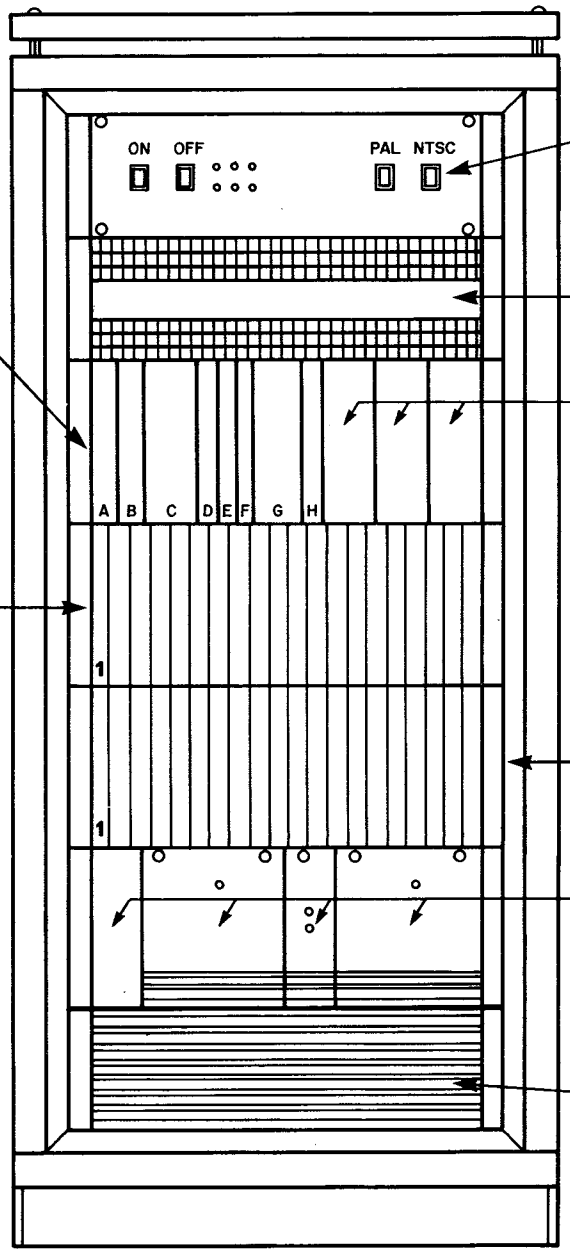
Other racks are concerned mainly with power supply and control, the amount of external control being confined to a minimum. Only the two controls ON/OFF and PAL/NTSC are provided, with provision for remotely controlling a relay which bypasses the machine if required in the event of an emergency. The relay also defaults to a bypass when the machine is switched off.

Compared with the three cabinets shown in Fig. 51, it can be seen that the size of the equipment has been considerably reduced.

9. The second field trial

To test the efficacy of the modifications just described the re-engineered equipment was installed in Television Centre for a two-week field trial in September 1978. A representative selection of programmes was chosen to be processed at the time of transmission and, as before, the pro-

A	Input processor
B	D.A.C.
C	Spares
D	
F	
G	Preset controls
H	Burst locked clock generator



1	Clock gen option and housekeeping
2	Input subtractor O.B. → S.M. converter
3	Sync. bypass L.D.
4	Sign L.D.
5	Filter L.D.
6	Filter L.D.
7	Filter L.D.
8	Filter L.D.
9	Five term summer and accumulator
10	Horizontal filter
11	Side-chain mult' and trans' function
12	S.M. → O.B. converter and com' delay.
13	Main path multiplier
14	Adder and error feedback
15	Grey-scale detector
16	Line delay
17	Line delay
18	Field store 1
19	Field store 2
20	Field store control

1	Line delay
2	Line delay
3	Band-pass filter feed
4	Transversal filter modules
5	Transversal filter modules
6	Transversal filter modules
7	Transversal filter modules
8	Transversal filter modules
9	Transversal filter modules
10	Transversal filter modules
11	Adder and band-pass filter control
12	Measurement unit 4
13	Measurement unit 3
14	Measurement unit 2
15	Measurement unit 1
16	Automatic control 1
17	Automatic control 2
18	Automatic control 3
19	Line delay
20	Line delay

Fig. 63 – Mechanical layout of the prototype noise reducer.

gramme signals were available to the appropriate Network Control, either processed or unprocessed. In this way an average of four programmes per night was processed during most of the trial period. In addition the whole of the Saturday afternoon sports programme on 16th September was processed and, on the evening of 19th September, the entire output of Network 1 was processed from 1700 hours onwards.

Questionnaires were completed for the whole period by staff in the Standards Conversion Area and by Quality Check when this was manned. In both cases input and output pictures were available for comparison. Observers were asked to judge the noise, before and after processing, and the movement degradation using the CCIR impairment scale and the overall effect using the CCIR comparative scale.

The observations during the first half of the period were subsequently discounted as the equipment was found to be operating incorrectly. After this had been remedied the average result was that the noise reducer reduced the visibility of noise by one grade from Grade 3.7 to Grade 4.7. In the case of 16 mm film, grain was judged to be reduced considerably by up to $1\frac{1}{2}$ grades. Movement degradation was rarely observed, with the exception of sports programmes involving constant rapid panning of cameras. In this case "fairly noticeable smearing on cuts to rapid pans" was recorded. Overall improvement to pictures, taking into account the noise reduction advantage at the expense of movement degradation, if any, was +1.7 on the comparison scale.

Operation on 525/60 NTSC was checked on the regular news exchange to ABC (USA) after conversion to 525 lines. This was effective, the outgoing pictures being quieter than the incoming 625-line pictures.

In addition to these regular tests two further experiments were conducted. At the request of the Television Service a low-light test was arranged at Television Centre. The light level was reduced from the then normal 1500 lux to approximately 700 lux, at which point noise was easily visible in saturated colours. Noise reduction reduced it to a level at least as low as with the normal lighting level.

Secondly, the effect of cascading noise reducers was investigated using a pair of video tape machines. This was not a conclusive test since insufficient material was examined, but it indicated that three reductions in tandem would increase movement degradation, but not necessarily to the point where it became unacceptable. This aspect of noise reduction requires further investigation in the light of subsequent improvements to the movement detector system.

Overall, it was clear that the modifications to the original equipment had fulfilled their primary objectives. However, three obvious shortcomings were observed as a result of using the equipment on such a wide range of programmes.

1. Low-frequency noise in some low-light areas of high contrast scenes was not always adequately dealt with.

2. Sports items, such as athletics and football, showed temporary smearing following cuts to panning shots. This smearing was very obvious and sometimes annoying.

3. The equipment was unduly disturbed by non-sync cuts and by any small disturbances to sync. pulses such as noise and jitter. The effect on the picture took the form of a temporary lateral shift of low-level picture information lasting about 1 second. This was infrequent and not very disturbing but obviously undesirable.

The first problem is a fundamental one, caused by the dependence of the noise measurement on the spectrum of the noise, as mentioned in section 7.3.1. It is also connected with the higher subjective visibility of low frequency noise.

The second problem was caused by the adaptive system trying to make a measurement of noise during global motion because of insufficient global motion sensitivity. The third problem was caused by inadequate provision, in the first clock generator, for dealing with non-sync cuts.

In spite of these shortcomings, the majority of programmes benefitted by noise reduction although by an inconsistent amount. The use of the equipment on the BBC1 network output for an entire evening illustrated the feasibility of using noise reduction to improve the consistency of a network output.

10. The third field trial and associated modifications

After the second field trial the equipment was further modified in an attempt to overcome the shortcomings listed above. The sensitivity to global motion was improved by implementing the detector based on line integrals instead of field

integrals, as described in Section 7.3.2. The non-sync cut performance was improved by adding circuitry to distinguish between cuts and the effect of impulsive interference as described in Section 8.2. In addition, the precaution was taken to inhibit the noise reduction action whenever the clock generator became unlocked, thereby preventing the lateral shift caused by incorrect timing. Further, the improved non-sync cut performance afforded the opportunity of increasing the basic stability of the clock oscillator.

In an attempt to improve the consistency of noise reduction in some difficult situations the noise measurement logic was modified. This involved the procedure to be followed when one or more grey ranges become non-bona-fide, as described in Section 7.3.4. In addition, the grey range spacing was slightly adjusted so as to give more even differences in measured noise between adjacent ranges. Care was also taken to exclude a small range near black level as spurious measurements could be caused by black clipping.

The equipment, with these modifications, was then installed in Television Centre for a more extended field trial in August 1979. After an initial period of some six weeks during which time it was used on selected programmes on both networks, it was permanently installed in the output feed of Network 2 for about 15 months.

During the initial period, operational staff in the Standards Conversion and Quality Check areas were again asked to fill in questionnaires on the same pattern as before and the results of these questionnaires are shown in Fig. 64. In presenting the results the programme material has first been divided into the two categories of electronic and film origination since their characteristics of motion portrayal are quite different. Secondly, in the light of experience, the electronically originated material has been further divided into "sport" and the rest.

As Fig. 64 shows, the noise levels of the two kinds of electronic pictures are very similar with film pictures being half a grade worse. Noise reduction improves the consistency of the noise level to grade 4.5 giving an improvement of 1.2 grades to film, the difference between "after" and "before" in Fig. 64. On the other hand, the grading shows that movement impairment was hardly ever noticed, although loss of definition during camera panning, particularly of grass, was commented on in the sporting items. It must be concluded that the observers associated "movement degradation" more with isolated objects in the scene rather than

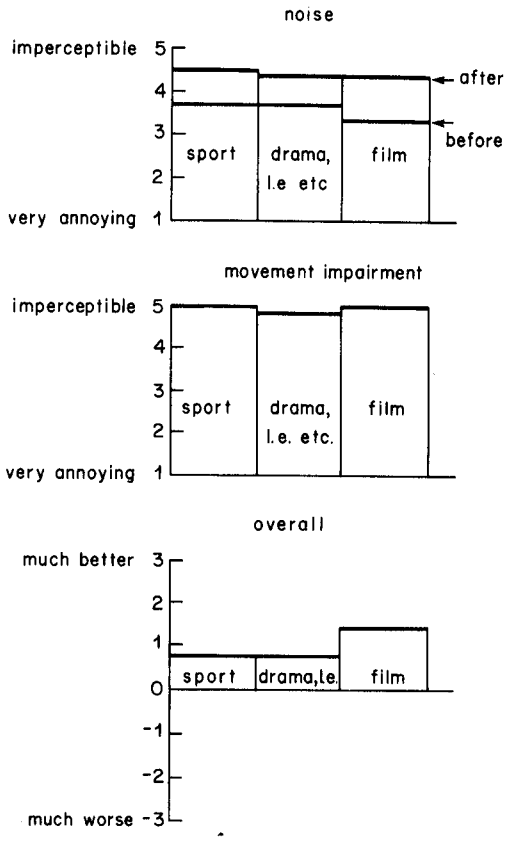


Fig. 64 – Results of the questionnaires in the third field trial.

with the background. Again, a slight loss of definition was recorded on several drama items and a few film items. Fine texture on facial skin presented an especial problem. Overall, however, the trade-off between noise reduction and movement effects was clearly thought to be worthwhile as judged by the overall ratings, film being improved by nearly $1\frac{1}{2}$ grades.

The results show that the modifications introduced before the trial largely achieved their object although sport, with its high proportion of cuts and panning texture, clearly presents a continuing difficulty. Most of it, however, was processed without serious impairment. The loss of definition in other electronically generated pictures could be an indication of the fundamental limit of sensitivity of the motion detector, given that pictures are rarely still. In addition, the removal of noise from a picture is well-known to reduce its subjective sharpness.

A further problem encountered, particularly with film where the noise levels tended to be higher, was that the noise was not removed consistently over the grey scale. This left "contours of noise" or, in the limit, some areas, corresponding to a

particular grey level, not noise reduced. This results from the coarse grey quantisation of the noise measurement process, coupled with measurement of the lowest noise level occurring within that range. In addition, at high noise levels, the "dirty window" effect, alluded to in the first field trial, was very apparent together with the re-appearance of noise at moving edges which was rather distracting. These effects are fundamental to the philosophy of noise reduction adopted in the equipment.

11. Further developments

The observations recorded by Television Centre staff during the third field trial indicate that the consensus was that an overall improvement in picture quality had resulted. In view of this, it was considered that further short term development of the system would be worthwhile to produce a general improvement in performance. Nevertheless, it was recognised that, in the longer term, a more complex approach could offer a fundamentally improved performance.

Before the third trial, work had already begun on two additional modifications. Firstly, a subcarrier locked clock generator was being developed, but was not ready at the time of the third trial. Secondly, a measure of image enhancement had been attempted which proved unsuccessful in its simple form.

11.1. Subcarrier-locked clock generator*

The basic stability of the clock oscillator, although adequate for most picture material, was inadequate for test signals. This was most noticeable on colour bars, where the residual jitter caused a slight random hue fluctuation, most visible in blue.

With a view to curing this problem a subcarrier-locked clock oscillator was developed. This was designed to have an instantaneous frequency, in the PAL case, of three times the subcarrier frequency, thereby being as close as possible to the original frequency of $851f_L$, but having phase perturbations to ensure vertical and temporal registration of sample points.

Fig. 65 shows the relationship between the sample sites corresponding to a constant-phase clock frequency of three times the subcarrier frequency and those corresponding to a phase-perturbed clock. Because the PAL subcarrier executes approximately one quarter of a cycle less than an integral number of cycles each line, the lines of constant subcarrier phase are shifted by approximately one quarter of a cycle each line. Thus the samples for a constant-phase, three times subcarrier clock may be regarded as having a shift of $\frac{1}{12}$ subcarrier cycle per line in the opposite direction, as shown. Vertically registered samples may, therefore, be obtained by selecting the appropriate pulses of a 12 times subcarrier clock, i.e. by dividing by four but skipping one pulse per line. After four lines the samples are coincident with those of an unperturbed, three-times subcarrier clock.

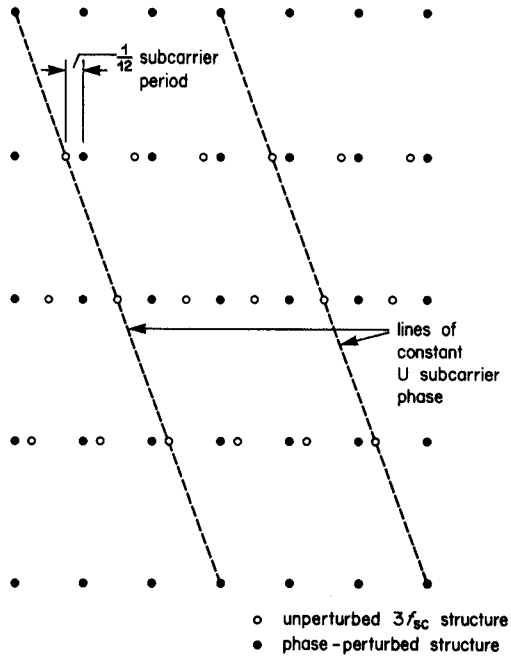


Fig. 65 – The relationship between sample sites of a phase-perturbed and unperturbed three times subcarrier clock.

The picture-frequency term in the definition of subcarrier frequency introduces a slight skew into the orthogonal structure so derived which accumulates to exactly one cycle or three samples after one picture. Thus if the line-by-line pulse skipping were to continue without interruption, picture-stable samples would be generated. However, a convention would then be needed to determine where to apportion the extra three samples. Alternatively, the three samples may be eliminated by further skipping 12 pulses per picture to give, in the case of System I, 851 samples on each line. If the extra skipping comprises six pulses at the end of each field, the sites on the interlaced field are also in vertical register with those on the prime field. Fig. 66 shows the relevant parts of the resulting clock waveform.

* The sub-carrier-locked clock generator was designed and commissioned by A. Roberts.

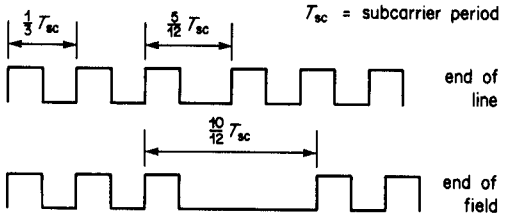


Fig. 66 – The waveform of the phase perturbed clock at line and field junctions.

For NTSC, the complications of phase perturbation can be avoided by choosing a clock frequency of four times the subcarrier frequency. This is only about $7\frac{1}{2}\%$ higher than the PAL clock frequency and is picture and line locked, with 910 samples per line.

Fig. 67 shows the block diagram of a clock generator embodying the above principles. Provision is made for a fall-back control system relying on the sync pulse information in cases where the subcarrier burst is absent. Further provisions are also made to deal with non-sync cuts.

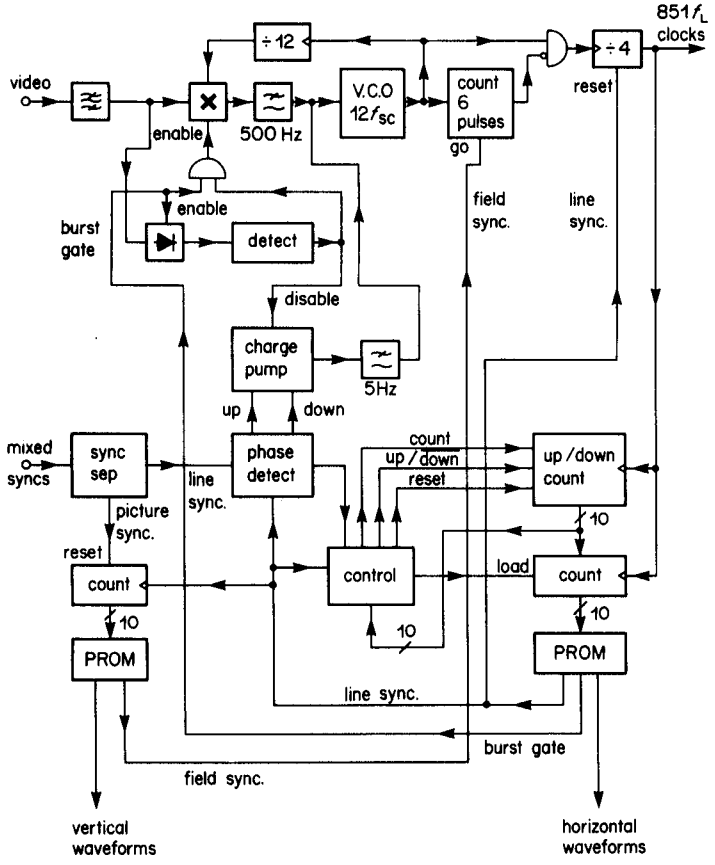


Fig. 67 – Block diagram of the phase-perturbed clock generator.

The voltage-controlled oscillator (VCO), running at $12f_{sc}$, is divided by 12 to obtain a subcarrier reference. During the burst gate this reference then multiplies the colour burst, obtained by bandpass filtering the video; the multiplier output, low pass filtered to nominally 500 Hz, controls the oscillator, subject to being enabled by the burst detector. If the burst detector indicates that the burst amplitude is too low, the oscillator is controlled by a charge pump signal, low pass filtered to nominally 5 Hz, derived from a line sync comparator, as formerly. Thus, the subcarrier-controlled loop is able to respond to legitimate timing fluctuations much more quickly than the line sync controlled loop.

The $12f_{sc}$ clocks are interrupted for six clocks every field on command of a suitably timed field pulse and then divided by four to obtain $3f_{sc}$ clocks. The divider is reset every line on command from a suitably timed line pulse to give 851 clocks per line at an instantaneous rate of $3f_{sc}$.

For NTSC, the clock pulse omission and divider resetting are inhibited and a second VCO running at $16f_{sc}$ is brought into play in conjunction with a divide-by-16 counter.

The output clocks are fed to a counter which addresses a PROM to produce the appropriate horizontal waveforms. One of these waveforms is a line pulse which is fed to a second counter to produce the vertical waveforms. In addition the line pulse is also used for resetting the clock divider and for line phase detection.

Provision for non-sync cuts is more elaborate than before. Detection and correction is performed by an auxiliary up/down counter which can operate in two modes in conjunction with a control circuit. In the quiescent detection mode it measures the interval between each incoming and locally-generated line sync. If this exceeds a limit the control circuit records whether the limit is exceeded for the next 16 lines. If it is, the last value is used to provide a coarse phase correction for the main counter. Thereafter the up/down counter switches to the measurement mode and integrates the phase error over the next 256 lines. This is used as the basis of a fine correction which is then implemented after which the up/down counter reverts to the detection mode. In this way cuts are detected and corrected more quickly, thereby minimising their effect on the higher bandwidth control loop of the subcarrier-locked oscillator. This clock generator effectively eliminated the jitter problem with no perceptible effects on saturated coloured areas.

11.2. Image enhancement

When pictures are enhanced by boosting the high frequencies the noise level is also increased. Often the limit to enhancement is set by the level of noise which is acceptable rather than the deviation from flatness of the frequency characteristic. It could be argued, therefore, that a reduction in noise could be exploited to provide further enhancement beyond that normally found in practice.

Enhancement is normally performed by horizontal and vertical aperture correctors which compensate for the aperture loss caused by the scanning spot. These act as transversal filters taking contributions from pixels adjacent to that being processed. A simple first-order corrector has a characteristic which peaks at half the reciprocal of the pixel spacing. This is normally chosen, horizontally, to give a peak beyond the cut-off frequency of the video signal, so being a fair match to the aperture loss. Vertically, however, there is no choice since the pixel spacing is the line pitch of a field. This gives a peak at the equivalent of 3.7 MHz for System I which is well within the video bandwidth and, if carried to excess, produces a characteristic, unnatural effect.

If, however, it is possible to take contributions from the lines of the interlaced field the pixel spacing is halved and the peak frequency is, correspondingly, doubled, giving a better match to the aperture loss. This is normally uneconomic since a field delay is required to gain access to the interlaced lines. The noise reducer configuration, however, provides a means of gaining access to the output signal delayed by up to, nominally, two fields and so, with little extra complication, out-of-band vertical correction may be implemented.

Accordingly, it was thought desirable to investigate whether such correction would give a worthwhile enhancement to normal pictures. A complication arises because such a corrector also affects moving pictures as the correction is derived from points separated in time as well as space. Analysis shows that the spatial high frequency boost is accompanied by a temporal high frequency boost of equal magnitude, peaking at the picture frequency. Whether or not this causes a problem depends on the ratio of the spatial to temporal loss in the source aperture and it was recognised that this could only be determined by implementing the corrector.

Fig. 68 shows a block diagram of the enhancement circuit actually used. The effect of the circuit is to add to the current line a factor k of it and

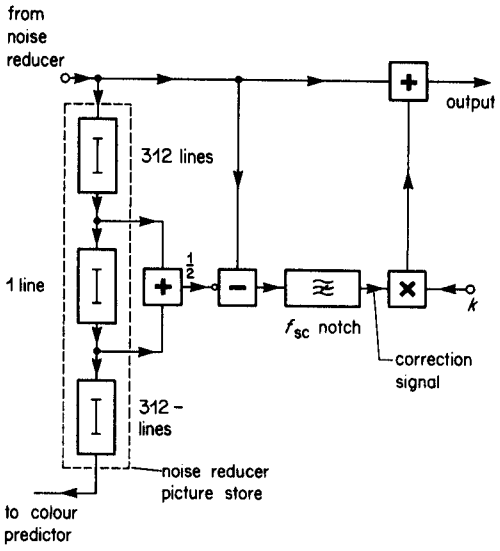


Fig. 68 – Block diagram of the noise reducer modification to give vertical aperture correction based on the lines of a picture.

subtract from it a factor $\frac{1}{2}k$ of each of the neighbouring lines in the previous field. A subcarrier notch filter is needed to ensure that the correction does not take place over the subcarrier region as the correction signal is completely inappropriate in that spectral region owing to the phase of the subcarrier.

In practice it was found that the corrector gave a visible improvement in vertical resolution on stationary pictures that were already relatively sharp. The minimum value of k for such pictures was found to be about 2. On unsharp pictures the value of k had to be increased considerably to give a noticeable improvement.

Unfortunately, however, the improvement in resolution that could be achieved before the noise level became objectionable was disappointingly low, especially with relatively unsharp pictures. More disturbing, however, were the temporal effects caused by the high frequency boost. Moving edges were surrounded by a "halo" increasing in visibility with the speed of motion. During rapid motion, the signal could exceed its permissible range causing problems with sync separators. Shot changes caused a similar overload lasting for a field.

With film, unsteadiness and twin-lens flying-spot telecine flicker caused by amplitude or registration imbalance were all exaggerated, giving the appearance of a badly misaligned, twin-lens telecine. In addition, the 25 Hz judder associated with film motion was accentuated.

Fig. 69 – The Compact Noise Reduction System designed by Pye TVT Ltd.

These effects rendered the corrector completely unusable for all but stationary electronic camera pictures. Nevertheless, if the benefits of out-of-band vertical correction were thought desirable, some of the undesirable temporal effects could be mitigated. For example the enhancement could, like the noise reduction, be adaptively controlled in response to motion. The accentuation of film motion judder could be avoided by deriving the correction signal alternately from previous and following fields. There is, therefore, scope here for further development.

12. Conclusions

A video noise reducer has been developed, based on a first-order recursive filter that uses a picture delay. Motion smearing is avoided by detecting motion and inhibiting the recursion wherever motion occurs, thereby bringing back the noise. The motion detector looks for variations in the power of the picture difference signal, allowing for the noise level. An automatic system has been developed for measuring the noise level which takes into account its variation over the grey scale divided into four segments.

Field trials of a prototype machine, handling the composite PAL signal at the network output of Television Centre, have confirmed the acceptability of this method of noise reduction for the vast majority of programme material containing moderate amounts of noise. Residual problems occur with material which has an excessively high level of noise or which contains extreme amounts of motion such as occurs in some sporting events but, on balance, they are outweighed by the overall improvement in subjective picture quality.

The design of the prototype has been sold to a commercial manufacturer* who has designed and manufactured, under licence, compact equipment shown in Fig. 69. This is currently in use on both BBC television networks.

13. References

1. COMANDINI, 1977. Signal processing in the Image Transform System. *SMPTE Journal*, **86,** August 1977, pp. 547–549. See also British Patent Specification 1402 609. Noise Reduction System for Video Signals.

2. TANTON, N.E., 1976. Video Digital Filter Study II: Temporal first-order recursive digital filter. BBC Research Department Report No. 1976/26.

3. KHADAVI, K. *and* ROGEL, P., 1978. Réduction de visibilité de bruit sur images de Télévision. Revue de radiodiffusion-télévision No. 52, 1978, pp. 35–42.

4. WALKER, R. *and* McNALLY, G.W., 1977. An experimental digital picture store. BBC Research Department Report No. 1977/9.

5. SANDERS, J.R. *and* WESTON, M., 1973. An automatic shot-change detector for telecine. BBC Research Department Report No. 1973/10.

6. British Patent Application No. 7926778, US Patent No. 4364087. Movement Detector for Television Signals.

* Pye TVT Ltd.

Appendix 1

The mean and variance of a rectified Gaussian distribution

The rectified Gaussian distribution of Fig. 27(a) is given by the expression:

$$\begin{aligned} p(x) &= 2(\sigma^2 2\pi)^{-1/2} \exp - x^2/2\sigma^2, \quad x > 0 \\ &= 0 \qquad\qquad\qquad\qquad\qquad\quad , \quad x < 0 \end{aligned}$$

The expectation of the distribution, $E(x)$, is given by:

$$\begin{aligned} E(x) &= \int_0^\infty xp(x)\,dx \\ &= 2(\sigma^2 2\pi)^{-1/2} \int_0^\infty x \exp(-x^2/2\sigma^2)\,dx \\ &= \sigma(2/\pi)^{1/2} \int_0^\infty y \exp(-y^2/2)\,dy \\ &= \sigma(2/\pi)^{1/2} \end{aligned}$$

The variance of the distribution, $v(x)$, is given by:

$$v(x) = E(x^2) - [E(x)]^2$$

$$\begin{aligned} \text{and } E(x^2) &= \int_0^\infty x^2 p(x)\,dx \\ &= \int_{-\infty}^\infty x^2 \tfrac{1}{2}[p(x) + p(-x)]\,dx \\ &= \sigma^2 \text{ by definition of the Gaussian function} \end{aligned}$$

$$\therefore v(x) = \sigma^2(1 - 2/\pi)$$

If the Gaussian distribution is biased before rectification by the amount V to give a distribution as in Fig. 29 then the distribution after rectification is the superposition of two distributions $p_1(x)$ and $p_2(x)$ where:

$$\begin{aligned} p_1(x) &= (\sigma^2 2\pi)^{-1/2} \exp[-(x - V)^2/2\sigma^2], \quad x > 0 \\ &= 0 \qquad\qquad\qquad\qquad\qquad\quad , \quad x < 0 \end{aligned}$$

and

$$\begin{aligned} p_2(x) &= (\sigma^2 2\pi)^{-1/2} \exp[-(x + V)^2/2\sigma^2], \quad x > 0 \\ &= 0 \qquad\qquad\qquad\qquad\qquad\quad , \quad x < 0 \end{aligned}$$

The expectation of the distribution, $E(x)$, is given by:

$$E(x) = \int_0^\infty x[p_1(x) + p_2(x)]\,dx$$

Putting $(x - V)/\sigma = y$ we have

$$\int_0^\infty xp_1(x)\,dx = (\sigma^2 2\pi)^{-1/2} \int_{-V/\sigma}^\infty \sigma(\sigma y + V) \exp(-y^2/2)\,dy$$

$$= \sigma(2\pi)^{-1/2}\exp(-V^2/2\sigma^2) + (2\pi)^{-1/2}V(\pi/2)^{1/2}[1 + \operatorname{erf}(V/\sigma\sqrt{2})]$$

$$= a(2\pi)^{-1/2}\sigma + \tfrac{1}{2}(1 + b)V$$

where

$$a = \exp(-V^2/2\sigma^2), \qquad b = \operatorname{erf}(V/\sigma\sqrt{2}) \qquad \text{and} \qquad \operatorname{erf}(x) = 2/\pi^{1/2}\int_0^x \exp(-y^2)\,dy$$

Substituting V for $-V$ we have:

$$\int_0^\infty xp_2(x)\,dx = a(2\pi)^{-1/2}\sigma + \tfrac{1}{2}(b - 1)V$$

since $$\operatorname{erf}(-x) = -\operatorname{erf}(x)$$

Thus the total expectation is:

$$E(x) = 2a(2\pi)^{-1/2}\sigma + bV = aE_0 + bV$$

where $E_0 = \sigma(2/\pi)^{1/2}$, the expectation of the unbiased distribution.

Appendix 2

The mean and variance of a root mean square of samples taken from a normal distribution

A root mean square of n samples is defined as:

$$R_n = n^{-1/2}(x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}$$

$$= n^{-1/2}\left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$$

where x_i is a sample with a certain probability distribution function.

The expectation of R_n is given by:

$$E_n \equiv E(R_n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} R_n p(x_1, x_2, \ldots, x_n)\,dx_1\,dx_2\ldots dx_n \qquad (1)$$

and if the samples x_i are independent samples from a normal distribution with zero mean and variance σ^2

$$p(x_1, x_2, \ldots, x_n) = p_1(x_1)p_2(x_2)\ldots p_n(x_n)$$

$$= (\sigma^2 2\pi)^{-n/2}\exp - (x_1^2 + x_2^2 + \cdots + x_n^2)/2\sigma^2$$

The integral in Equation (1) is a hypervolume integral in a space of n dimensions. Letting

$$\sum_{i=1}^{n} x_i^2 = r^2$$

where r is the radius of a hypersphere of n dimensions, Equation (1) may be written:

$$E_n = \int_0^\infty n^{-1/2}r(\sigma^2 2\pi)^{-n/2}\exp - r^2/2\sigma^2\,dV_n$$

where the hypersphere volume V_n is defined as:

$$V_n = k_n r^n$$

Substituting for V_n and putting $y = r/\sigma$ we have:

$$E_n = n^{1/2} k_n \sigma (2\pi)^{-n/2} I_n$$

where

$$I_n = \int_0^\infty y^n \exp(-y^2/2)\,dy$$

Now, integrating by parts, it can be shown that:

$$I_n = (n-1) I_{n-2}$$

with

$$I_0 = (\pi/2)^{1/2} \quad \text{and} \quad I_1 = 1$$

Thus

$$I_{2n} = [(2n)!/(n!2^n)](\pi/2)^{1/2}$$

and

$$I_{2n+1} = 2^n n!$$

Therefore

$$E_{2n}/\sigma = [(2n)!/(n!2^{2n})](n\pi)^{1/2}\pi^{-n}k_{2n}$$

and

$$E_{2n+1}/\sigma = n!(2n+1)^{1/2}(2\pi)^{-1/2}\pi^{-n}k_{2n+1}$$

Now the volume of a hypersphere, and hence k_n, may be found by induction. Considering the derivation of the volume of a sphere from the "volume" of a circle, we have:

$$V_{n+1} = \int_{-r}^{r} V_n(y)\,dx$$

where $y^2 = r^2 - x^2$. But $x = r\cos\theta$ and $y = r\sin\theta$

$$\therefore\ V_{n+1} = 2\int_{\pi/2}^{0} -V_n(r\sin\theta)\, r\sin\theta\, d\theta$$

and putting $V_{n+1} = k_{n+1} r^{n+1}$ and $V_n = k_n y^n$ we have:

$$k_{n+1} r^{n+1} = 2\int_0^{\pi/2} k_n (r\sin\theta)^n r\sin\theta\, d\theta$$

whence

$$k_{n+1} = 2k_n I_{n+1}$$

where

$$I_n = \int_0^{\pi/2} \sin^n\theta\, d\theta$$

Integration by parts shows that $I_n = n^{-1}(n-1)I_{n-2}$ with $I_0 = \pi/2$ and $I_1 = 1$

so that

$$I_{2n} = (2n)!(n!2^n)^{-2}(\pi/2)$$

and

$$I_{2n+1} = (n!2^n)^2/(2n+1)!$$

Now

$$V_1 = 2r \quad \text{and} \quad V_2 = \pi r^2$$

whence

$$k_1 = 2 \quad \text{and} \quad k_2 = \pi$$

Thus

$$\begin{aligned} k_{2n} &= 2I_{2n}k_{2n-1} \\ &= 2^2 I_{2n}I_{2n-1}k_{2n-2} \\ &= (\pi/n)k_{2n-2} \\ &= \pi^n/n! \end{aligned}$$

and

$$\begin{aligned} k_{2n+1} &= 2I_{2n+1}k_{2n} \\ &= 2^2 I_{2n+1}I_{2n}k_{2n-1} \\ &= 2\pi(2n+1)^{-1}k_{2n-1} \\ &= 2(2\pi)^n n!2^n/(2n+1)! \end{aligned}$$

Thus

$$E_{2n}/\sigma = a(2n)^{1/2}$$

and

$$E_{2n+1}/\sigma = a^{-1}(2n+1)^{-1/2}$$

where

$$a = (\pi/2)^{1/2}(2n)!/(n!2^n)^2$$

The variance of R_n is given by:

$$v_n \equiv v(R_n) = E(R_n^2) - [E(R_n)]^2$$

Now

$$E(R_n^2) = E\left(n^{-1}\sum_{i=1}^{n} x_i^2\right)$$

$$= n^{-1}\sum_{i=1}^{n} E(x_i^2)$$

But $E(x_i^2) = \sigma^2$ if the samples x_i are taken from a normal distribution with zero mean.

$$\therefore E(R_n^2) = \sigma^2$$

and

$$v_n/\sigma^2 = 1 - (E_n/\sigma)^2$$

Appendix 3

The probability distribution of a minimum of a number of uncorrelated samples taken from a normal distribution

Let $p(x)$ be the probability distribution function of the individual uncorrelated samples and $p_n(x)$ be the probability distribution of the minimum of n samples.

A given value of the minimum may occur if any of the samples takes that value irrespective of the other sample values provided they are equal to or greater than it.

The probability that any given sample takes the minimum value is:

$$p(x)\,\mathrm{d}x\left[\int_x^\infty p(x)\,\mathrm{d}x\right]^{n-1}$$

This may happen for the n different samples so that the minimum may occur n different ways. Thus the probability of the minimum is

$$p_n(x)\,\mathrm{d}x = np(x)\,\mathrm{d}x\left[\int_x^\infty p(x)\,\mathrm{d}x\right]^{n-1}$$

If the samples are taken from a normal distribution then

$$p(x) = (\sigma^2 2\pi)^{-1/2} \exp\left[-(x-\mu)^2/2\sigma^2\right]$$

where μ and σ^2 are the mean and variance of the distribution.

Putting $y = (x-\mu)/\sigma$ we have

$$\int_x^\infty p(x)\,dx = (\sigma^2 2\pi)^{-1/2} \int_y^\infty \sigma \exp(-y^2/2)\,dy$$
$$= \tfrac{1}{2}[1 - \operatorname{erf}(y/\sqrt{2})]$$

Thus $p_n(x) = n(\sigma^2 2\pi)^{-1/2} 2^{1-n}[1 - \operatorname{erf}(y/\sqrt{2})]^{n-1} \exp(-y^2/2)$ with $y = (x-\mu)/\sigma$.